TURING 图灵新知



Unknown Quantity:
A Real and Imaginary History of Algebra

代数的历史

人类对未知量的不舍追踪

【美】 John Derbyshire 著

冯速 译





人民邮电出版社
POSTS & TELECOM PRESS

Unknown Quantity:
A Real and Imaginary History of Algebra

代数的历史

人类对未知量的不舍追踪

美国《图书馆杂志》2003 年度最畅销科普著作！

美国《基督教科学箴言报》2003 年度非小说类最值得关注图书！

《科学图书和电影杂志》2003 年度最佳图书！

生活在四千年前的古巴比伦人的成就缘何可以与文艺复兴时期的意大利相媲美？丢番图和花拉子米到底谁才是真正的代数之父？虚数是历经了怎样的磨难才被人广为接受的？牛顿和高斯的伟大体现在何处？旷世奇才格罗申迪克是如何书写他的传奇人生的？

来吧，走进《代数的历史》，和 Derbyshire 一起穿过历史迷雾，体味代数这门最纯净、最严苛的智力学科之非凡魅力，揭开未知量 x 的前世今生，探寻现实世界最深层、最本质的秘密！

John Derbyshire 知名科普作家，数学家，教育语言学家，专业系统分析师。除本书外，他还著有广受好评的 *Prime Obsession* 和 *Seeing Calvin Coolidge in a Dream*。他的作品经常刊登在美国《国家评论》和《新标准》等杂志上。





封面设计：于洋

图灵网站：www.turingbook.com　热线：(010) 51095186

反馈 / 投稿 / 推荐信箱：contact@turingbook.com

有奖勘误：debug@turingbook.com

分类建议　科普读物/数学

人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-22537-5

定价：35.00元

TURING 图灵新知



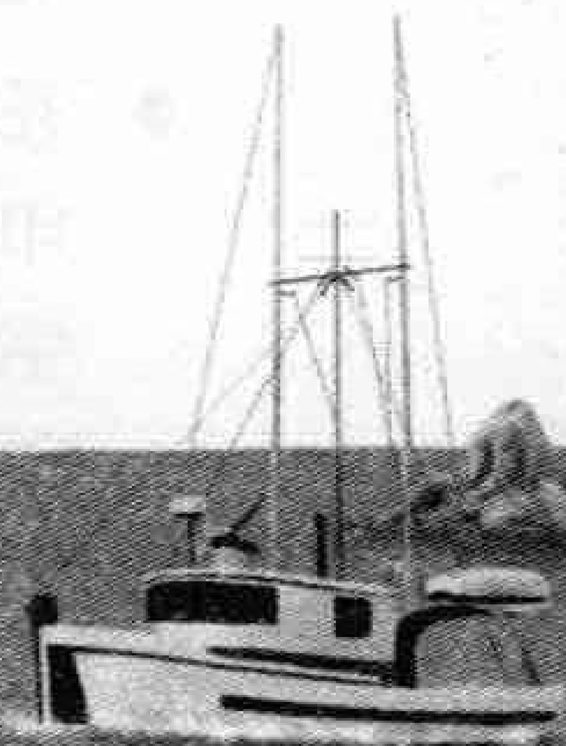
Unknown Quantity:
A Real and Imaginary History of Algebra

代数的历史

人类对未知量的不舍追踪

【美】 John Derbyshire 著

冯速 译



人民邮电出版社

北 京

图书在版编目（C I P）数据

代数的历史 ：人类对未知量的不舍追踪 /（美）德比希（Derbyshire, J.）著 ；冯速译. -- 北京 ：人民邮电出版社，2010.8
（图灵新知）
书名原文：Unknown Quantity：A Real and Imaginary History of Algebra
ISBN 978-7-115-22537-5

Ⅰ. ①代… Ⅱ. ①德… ②冯… Ⅲ. ①代数—普及读物 Ⅳ. ①015-49

中国版本图书馆CIP数据核字(2010)第043609号

内 容 提 要

本书是一本介绍代数发展历史的科学普及读物，作者以轻松诙谐的笔触将代数几千年来的重大事件和重要人物展现出来，让读者从一个侧面对整个数学的发展有总体的认识。

本书适合中学生至大学生等各层次的数学爱好者阅读，也是研究数学史极有价值的参考读物。

图灵新知

代数的历史：人类对未知量的不舍追踪

◆ 著 [美] John Derbyshire
译 冯 速
责任编辑 朱 巍

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子函件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
北京隆昌伟业印刷有限公司印刷

◆ 开本：880×1230 1/32
印张：9.625
字数：258千字 2010年8月第1版
印数：1－4 000册 2010年8月北京第1次印刷

著作权合同登记号 图字：01-2009-0823号

ISBN 978-7-115-22537-5

定价：35.00元

读者服务热线：(010)51095186 印装质量热线：(010)67129223
反盗版热线：(010)67171154

版 权 声 明

引　言

本书讲述代数的历史，为那些有强烈好奇心的非数学专业人士而写。作为这样一本书的作者，我似乎应该从告诉读者什么是代数开始。那么什么是代数呢？

前几天我在一家机场书店逛，发现那里展示了高中生和本科生常用的公式表小折子[1]。公式表上有某一学科的所有基础知识，这些内容印在可以折叠的一联三幅的透明塑料上。其中有两个小折子是关于代数的，标题分别是“代数　部分1”和“代数　部分2”。副标题上写着，这两部分“涵盖了小学、中学和大学课程中的数学原理”。

我浏览了其中的内容。有些话题在专业数学人士看来并不是代数话题。例如，“函数”、“数列和级数”应该属于数学家们称为“分析”的范畴。但是总体来说，这两个小折子概括了基础代数的主要内容，并明确给出了现行美国高中和本科生基础课程中“代数”一词的惯用定义：代数是高等数学的一部分，不属于微积分。

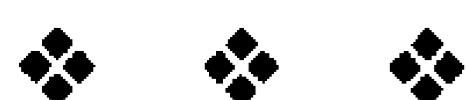
然而，在高等数学中，代数作为一门独立的学科有其鲜明的特性。20世纪伟大的德国数学家赫曼·韦尔曾在他1939年发表的一篇文章中给出了一句非常著名的评语[2]：

> 最近，拓扑天使和抽象代数恶魔正在为争取各个数学领域的幽灵而战。

读者也许知道拓扑是几何的一个分支，有时候称它为“橡胶几何”，它研究的是在拉伸、挤压但不撕裂的情况下图形所保持的性质。（对此不了解的读者可以先看第14章关于拓扑的详尽介绍，也可以参见韦尔的文

章对这一内容的更多介绍。）拓扑告诉我们，平常的环与打结的环之间的差异，球表面与甜甜圈表面之间的差异。为什么韦尔要把几何研究和代数对立起来呢？

也可以参见第15章开始给出的那份获奖清单，其中列出了近年来柯尔代数奖的获奖情况。非分歧类域论、雅可比簇、函数域、原动力上同调……显然，我们已经偏离四次方程和绘图很远了。正常的路线是什么呢？最简洁的答案就隐含在韦尔的评语之中：抽象。



当然，整个数学就是抽象的。几千年前人类就发现了数值，这就是数学抽象的最早活动。他们完成了从3个手指、3头牛、3个兄弟、3颗星星这样观察到的3的实例，到自身赋有含义的精神对象3的想象飞跃，这个3无需再表示为3个手指之类的特殊实例。

第二次这样的活动是抽象层次的第二次提升，标志是公元1600年前后几十年间被逐步采用的文字符号体系，即使用字母符号来表示任意数即未知数：数据（给定的事物）或者目标（要求解的事物）。艾萨克·牛顿称其为“泛算术”。能够走过这一漫长而充满羁绊的旅程，足以证明当时的人们渴望求解方程，或者说渴望确定某些数学情形下的未知量。这是一次在人类大脑中播下“代数”种子的旅程，也是我在本书的第一部分将要介绍的核心内容。

如果有人问公元1800年的一位受过良好教育的人代数是什么，他也许会说代数就是在计算或求解方程过程中使用字母符号来“放飞想象”（莱布尼茨）。在当时的欧洲，掌握或者至少了解数学文字符号的用法是教育的一个要求。

然而在19世纪[3]，这些字母符号开始从数值领域分离出来。人们发现[4]了各种奇怪的新数学对象[5]：群、矩阵、流形以及其他很多对象。数学开始飞向新的抽象层次。这是采用文字符号体系后一个自然而然的过程，它表明符号系统已经完完全全深入人心了。因此，把这种分离看成

是代数历史的延续不无道理。

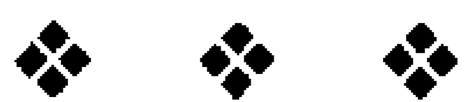
为此，我将把本书分成如下三个部分。

第一部分：从人类发现数值到大约公元1600年文字符号体系（即用字母表示数值）被广泛使用。

第二部分：从符号体系在数学领域取得辉煌战果到传统算术和几何概念中的符号缓慢分化，最终导致了新数学对象的发现。

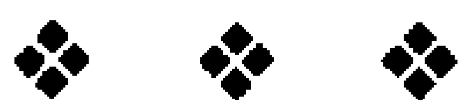
第三部分：把新数学对象置于坚实的逻辑基础之上、为更高层次的抽象铺平道路的近世代数。

代数的发展与所有人类活动一样，是不规则且无计划性的，我很难严格按照年代顺序陈述，特别是19世纪的代数。然而，我希望我的陈述方式被大家认可，能让读者对代数的主要发展线索有清晰的认识。



我的目的不是向读者传授高等代数。有很多这方面的优秀教科书，在叙述过程中我也会推荐一些这样的书。我的这本书不是教科书。我只希望能够大致展示代数中的一些概念，以及后来的代数是如何从早期代数发展而来的，哪些人扮演了重要的角色，历史环境又是怎样的。

然而，我发现不稍微解释一下这些代数学家都做了什么，是不可能说清楚这一学科的历史的，于是就有了书中一定量的数学解说。对于那些高中课本中通常不会讲到的内容，我把它们简单地归纳了一下，放在“数学知识”部分。这部分内容你有必要通读才能跟得上正文的历史陈述。每一部分数学知识都介绍若干基础概念。在某些地方，我会扩展正文中的概念，意在让已经学过某些本科数学课程的读者温故而知新，或为那些没有这样经历的读者提供最基本的知识。

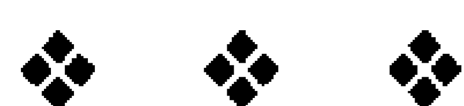


当然这本书是编著的，参考了很多其他人的著作。我将在正文和注

解里注明引用的作品。有三个资料来源我会经常提到，因此有必要在一开始就提醒自己不能忘了致谢。第一个资料来源是非常宝贵的《科学传记辞典》（*Dictionary of Scientific Biography*，后文中有时也简称DSB），它不仅列出了数学家的详细生平，而且还给出了数学思想的起源和数学思想传播的宝贵线索。

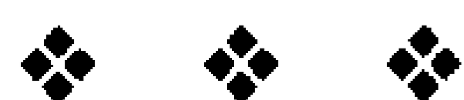
另外两本主要参考的著作是数学家为数学家所写的代数历史：范德瓦尔登的《代数历史》和伊莎贝拉·巴什科娃与加里娜·史莫诺瓦合著的《代数的起源与演变》（2000年由阿贝·申尼兹尔译成英文）。在本书中我引用这些书时将直接引用其作者的名字（如“范德瓦尔登说……”）。

在这里还要感谢另外一位为本书做出重要贡献的人——芝加哥大学的理查德·斯旺教授。他审阅了本书的手稿，能得到他的指点我感到万分荣幸。斯旺教授对内容提出了很多意见、批评、修正和建议，大大提升了本书水准。我忠心感谢他的帮助和鼓励。尽管我力争做到完美，但书中仍然会存在一些错误或者疏忽，对此我负全责。



这是一本关于代数的故事书。它开始于遥远的过去，讲述从“这个加这个等于这个”的陈述到“这个加什么等于这个”的提问的简单思维转变。和代数打交道一定会遇到x这个未知量，它是第一个闯入人类思维的，过了很长时间后才有了表示未知量和任意数值的符号体系的需求。一旦建立起这样的符号体系，方程的研究就进入到了更高的抽象层次。其结果是，新的数学对象出现，导致向更高抽象层次的飞跃。

现在，代数已经成为所有智力学科中最纯净、最苛刻的学科，它的对象抽象再抽象，非数学人士几乎无法领会到其成果的巨大作用和非凡魅力。最令人惊讶，也最神秘的是，在这些非物质的精神对象层层嵌套的抽象之中，包含着物质世界最深层、最本质的秘密。



注 解

[1] 由位于博卡拉顿的BarCharts公司出版，其作者是S.B.基利克。

[2] 引自《杜克数学杂志》第5期的“不变量”一文。

[3] 有时候，也许我要像历史学家约翰·卢卡克斯那样使用“此19世纪”来表示1815年到1914年的那段时期。然而，这里还是指通常意义上的19世纪。

[4] 发现还是发明？我倾向于“柏拉图”的观点：这些对象已经在世界的某个地方存在，等待着人类去发现它们。这是大脑中的框架，由非常专业的数学家利用大量的时间构建了大部分的数学。这一点非常了不起，但是它与代数历史的关系不大，因此我不再赘述。

[5] “数学对象”指的是专业数学家感兴趣的一种东西，他们竭力去了解和开发关于它们的定理。非数学专业人士最熟悉的数学对象是数值和点、线、三角形、圆、立方体等，它们都是欧几里得几何的二维空间和三维空间中研究的对象。

目　　录

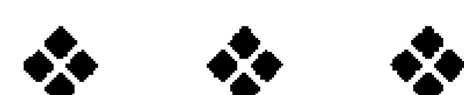
第三部分　抽 象 层 次

数学知识

数值和多项式

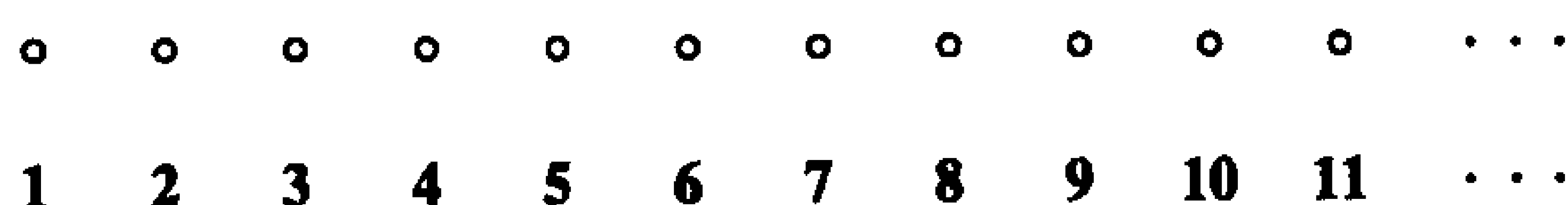
我将在书中适当的地方插入一些数学知识，简单地描述相应的历史，让你了解或提醒你注意一些数学概念，以便你能够顺利理解后续将要讲述的历史。

这是全书的第一组数学知识，其中包含两个你必须掌握的数学概念，这样才能够理解正文中的内容。这两个概念就是数值和多项式。



现代意义上的数值的概念在19世纪后期开始成型，并于20世纪20年代和30年代在数学界广泛传播。它就像层层嵌套的“俄罗斯娃娃”的模型，模型中共有五个“俄罗斯娃娃”，分别用镂空的字母表示为$\mathbb{N}$、$\mathbb{Z}$、$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$。

最里面的“娃娃”是自然数，所有自然数的总体记为$\mathbb{N}$。这些是普通的[1]计数数值，如1、2、3等。可以把它们形象地排列成向右无限延长的点线，如图NP-1所示。

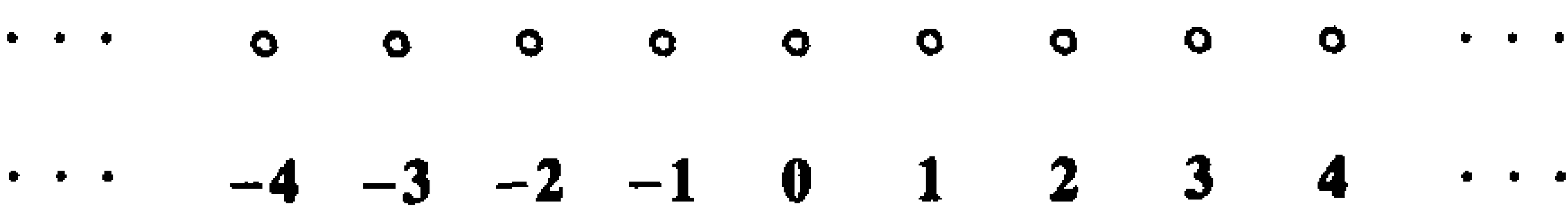


图NP-1 自然数$\mathbb{N}$的家族

自然数非常有用，但是有一些缺点。主要的缺点是从一个自然数中减去另一个自然数不是永远可行，用一个自然数除以另一个自然数也不是永远可行。你可以用7减去5，但是不能用7减去12——我的意思是说，

如果你想要得到一个自然数结果，这是不行的。专业术语就是：在减法下ℕ不是封闭的，在除法下ℕ也不是封闭的。你可以用12除以4，但不能用12除以5，因为这样结果就不是自然数了。

减法的问题因为零和负数的发现而得到了解决。大约公元600年，印度数学家发现了零。负数是欧洲文艺复兴时期的成果。扩张自然数系统使其包含这些新对象，就得到了第二个“俄罗斯娃娃”，它包含第一个“俄罗斯娃娃”。这个数系就是整数，整数集合用符号ℤ表示（来自于德语的“数”，Zahl）。整数可以形象地用向左右两端无限延长的点线表示，如图NP-2所示。



图NP-2　整数ℤ的家族

现在，我们可以随意进行加、减、乘运算，当然做乘法运算需要了解符号法则。

> 正正得正。正负得负。负正得负。负负得正。

更简单地说：相同符号相乘得正，不同符号相乘得负。当除法成立时，符号法则也可以运用于除法，如−12除以−3得4。

然而，在ℤ中除法不是总可行。在除法下ℤ不是封闭的。为了得到一个在除法下封闭的数系，我们还要再次扩张整数系统，引入分数（包括正分数和负分数）。这样就有了第三个“俄罗斯娃娃”，它包含前两个数系。这个“娃娃”称为有理数，有理数的集合记为ℚ（来自于英文单词quotient）。

有理数是“稠密的”。这表明在任意两个有理数之间，你总可以找到另外一个有理数。ℕ和ℤ都不具有这样的性质。11和12之间没有自然数，−107与−106之间也没有整数。然而，在有理数1190507/10292881和185015/1599602之间总可以找到一个有理数，虽然这两个有理数相差不到16万亿分之一。例如，有理数2300597/19890493比上面的第一个有理

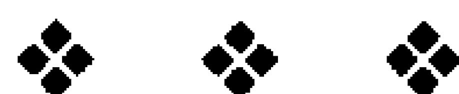
数大，但是比第二个小。很容易证明，因为任意两个有理数之间有其他有理数，所以你可以在任意两个有理数之间找到无限多个有理数。这就是有理数“稠密”的含义。

因为$\mathbb{Q}$具有稠密的性质，所以它可以用向左右两端无限延伸的连续直线表示。每一个有理数在这条直线上都有一个位置，见图NP-3。（注意：我们可以用相同的图形表示实数家族$\mathbb{R}$。）明白整数之间的空隙是如何被填充的吗？任意两个整数，比如说27和28，其间的有理数都是稠密的。



图NP-3　有理数$\mathbb{Q}$的家族

这些“俄罗斯娃娃”是嵌套的，$\mathbb{Q}$套着$\mathbb{Z}$，$\mathbb{Z}$套着$\mathbb{N}$。还有另外一种看待这一问题的方法：自然数是“名誉整数”，整数和自然数是“名誉有理数”。为了强调，可以把名誉数打扮成适当的模样。自然数12可以被装扮成整数+12，或者有理数12/1。



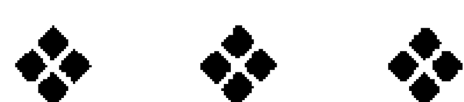
还有另外一些数值，它们既不是整数也不是有理数。大约公元前500年，希腊人发现了这种数。这一发现给希腊人的思维带来了深刻的影响，而且还引出了数学家和哲学家至今也没有给出满意答案的问题。

这样的数的最简单例子是2的平方根，如果你把它与它自己相乘，就得到数值2。（从几何观点看：边长是单位1的正方形的对角线的长度就是这个数。）很容易证明，没有有理数可以表示边长为单位1的正方形的对角线长度。[2]用极其类似的方法可以证明，如果N不是完全k次幂，那么N的k次方根一定不是有理数。

显然，我们需要另外一个“俄罗斯娃娃”，它要能够包含所有这些无理数。这个“新娃娃”就是实数，用$\mathbb{R}$表示。2的平方根是一个实数，但不是有理数：它在$\mathbb{R}$内但不在$\mathbb{Q}$内。（当然也不在$\mathbb{Z}$和$\mathbb{N}$内。）

同有理数一样，实数也是稠密的。在任意两个实数之间总能找到另外一个实数。因为有理数总是稠密的，总能“填满”图示线，你也许会提出这样的疑问：如何把实数挤进有理数之间？更加怪异的是，ℚ 和 ℤ 是可数的，而 ℝ 是不可数的。可数集合的意思是，其中的元素可以与计数数值集合 ℕ 中的元素相匹配：1，2，3，等等，可以一直到无穷。然而，对于实数 ℝ 你做不到这一点。从某种意义上讲，ℝ 太大，比 ℕ 、ℤ 和 ℚ 都大，因此不能像上面那样计数。那么，这样超级无穷多的实数怎么能够被安插在有理数之间呢？

这是一个非常有趣的问题，它甚至让数学家们伤透了脑筋。然而，这不在代数历史的范畴内，我在这里提到这个问题只因为在第14章要提到可数性的问题。记住以下这点就足够了：表示 ℝ 的图示和有理数的图示非常类似，都是一条向左右两边无限伸展的连续直线（见图NP-3）。当这条直线用于图示 ℝ 时，它被称为“实线”。更抽象地说，可以把“实线”作为 ℝ 的同义词。



在 ℕ 内，加法和乘法总可行，减法和除法有时候可行。在 ℤ 中，加法、减法和乘法总可行，除法有时可行。在 ℚ 内，加法、减法、乘法和除法（在数学中永远不能有除以0的除法）都永远可行，但是开方却引出了问题。

ℝ 可以解决这些问题，但只限于正数。根据符号法则，当任意数与自己相乘时，都得到一个正数。用另外一种方式说就是：在 ℝ 内，负数没有平方根。

进入16世纪，这一限制开始成为数学家前进的障碍，所以必须向这一数系中加入新的“俄罗斯娃娃”。这个新“娃娃”就是复数，记为 ℂ 。在其中，每一个数都有平方根。事实证明，只用普通的实数和一个新数 $\sqrt{-1}$（通常记为i）就可以构建这个新数系。例如-25的平方根是5i，因为$5i\times5i=25\times(-1)=-25$。那么i的平方根是什么呢？这不难回答。我们熟悉的乘开括号的法则是$(u+v)\times(x+y)=ux+uy+vx+vy$。所以有：

$$\left(\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}\mathrm{i}\right)\times\left(\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}\mathrm{i}\right)=\frac{1}{2}+\frac{1}{2}\mathrm{i}+\frac{1}{2}\mathrm{i}+\frac{1}{2}\mathrm{i}^2$$

而且，因为$\mathrm{i}^2=-1$，1/2+1/2=1，所以上面等式的右边正好等于i。因此，左边括号中的数值是i的平方根。

这些“俄罗斯娃娃”也是嵌套的。实数x是“名誉复数”$x+0\mathrm{i}$。（形如$0+y\mathrm{i}$或简写为$y\mathrm{i}$的复数称为虚数，其中的y为实数。）

因为$\mathrm{i}^2=-1$，很容易推出复数的加法、减法、乘法和除法法则，如下所示。

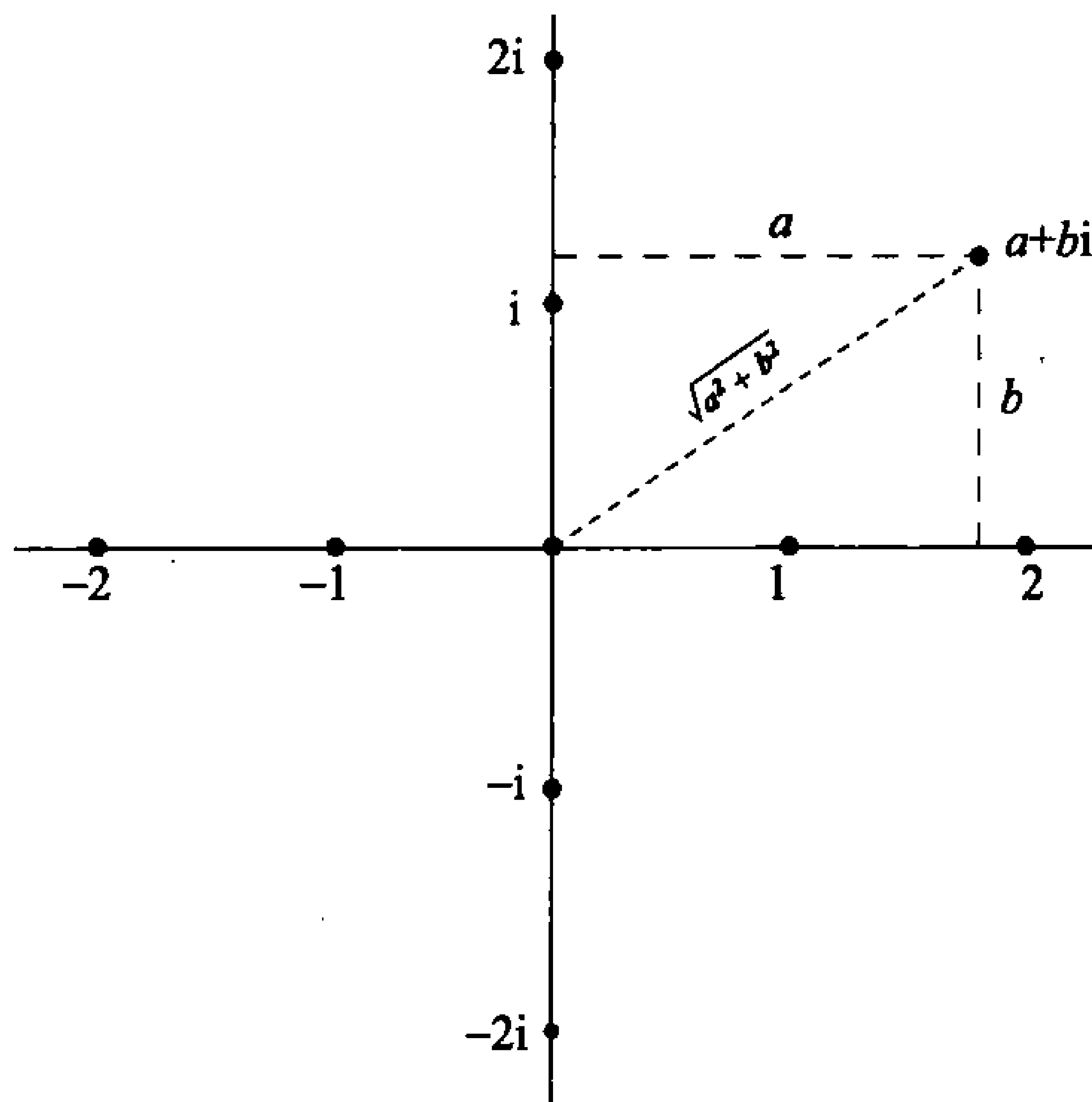
加法：$(a+b\mathrm{i})+(c+d\mathrm{i})=(a+c)+(b+d)\mathrm{i}$

减法：$(a+b\mathrm{i})-(c+d\mathrm{i})=(a-c)+(b-d)\mathrm{i}$

乘法：$(a+b\mathrm{i})\times(c+d\mathrm{i})=(ac-bd)+(ad+bc)\mathrm{i}$

除法：$(a+b\mathrm{i})\div(c+d\mathrm{i})=\dfrac{ac+bd}{c^2+d^2}+\dfrac{bc-ad}{c^2+d^2}\mathrm{i}$

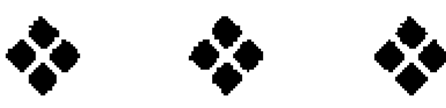
因为复数有两个独立的部分，所以$\mathbb{C}$的图示不能是直线。需要用一个向各个方向无限伸展的平面来表示$\mathbb{C}$。这个平面称为复平面（见图NP-4）。一个复数($a+b$i)可以用普通的上下、左右坐标表示为这个平面上的一个点。



图NP-4　复数$\mathbb{C}$的家族

注意，对于每一个复数($a+b$i)，都有一个非常重要的实数与之相关，这个实数称为该复数的模，记作$\sqrt{a^2+b^2}$。希望读者从图NP-4可以看出来，根据毕达哥拉斯定理[3]，一个复数的模就是在复平面内它到零点的距离，零点通常称为原点。

以后我们还会遇到其他数系，但是，所有这些数系都是以这五个依次嵌套的基础数系为基础的：$\mathbb{N}$、$\mathbb{Z}$、$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$。



有关数值就介绍到这里。本书经常提到的另一个关键概念就是多项式（polynomial）。这一单词的词源是希腊语和拉丁语的混合体，其意义是“有很多名字”，其中的“名字”可以理解为“命名部分”。似乎是法国数学家弗兰索瓦·韦达在16世纪后期首先开始使用这一名词的，100年后英国人才开始使用。

多项式是利用加、减和乘等运算把数和“未知量”联系起来而构建成的数学表达式（不是方程，这里没有等号）。这些运算可以出现任意多次，但不能是无限次。下面都是多项式的例子：

$$5x^{12}-22x^7-141x^6+x^3-19x^2-245$$

$$9x^2-13xy+y^2-14x-35y+18$$

$$2x-7$$

$$x$$

$$\frac{211}{372}x^4+\pi x^3-(7-8\mathrm{i})x^2+\sqrt{3}x$$

$$x^2+x+y^2+y+z^2+z+t^2+t$$

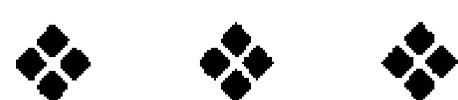
$$ax^2+bx+c$$

注意以下事项。

- 未知量。在一个多项式中可以有任意多个未知量。
- 使用字母表示未知量。一个真正的未知量，也就是我们真正想知道其值的量（拉丁语是quaesita，即“要求解的事物”），通常用拉丁语字母表中后面几个字母表示，x、y、z、t是最常用的表示未

知量的字母。

- **未知量的幂**。因为可以做任意有限次乘法，未知量的任意自然数次幂都可以表示出来，如x、x^2、x^3、x^2、y^3、x^5yz^2等。
- **使用字母表示“已知量”**。“已知量”的拉丁语是data，通常是取自于ℕ、ℤ、ℚ、ℝ或ℂ的数值。我们可以用表示已知量的字母作为多项式的系数。这些字母通常取自于字母表的开头（a、b、c等）或者中间（p、q、r等）。
- **系数**。现在，data作为一个英语单词有了它自己的含义，不能总说是“已知量”。多项式中的“已知量”现在称为**系数**。上面的第三个多项式的系数是2和−7。第四个多项式（严格地说它是一个单项式）的系数是1。最后一个多项式的系数是a、b和c。



多项式只是所有数学表达式中的一个小子集。如果把除法引入其中，可以得到更大的一类表达式，称为**有理表达式**，例如：

$$\frac{x^2-3y^2}{2xz}$$

这是一个有三个未知量的有理表达式，而不是多项式。引入更多的运算可以进一步扩大这个集合：开方，取正弦、余弦或对数，等等。最后所得到的表达式都不是多项式。

怎么写多项式。取某些“已知”数，可以是数值（17、$\sqrt{2}$、π等），也可以是代表数值的字母（a、b、c等，p、q、r等）。加入一些未知量（x、y、z等），进行有限次加、减和乘运算，其结果就是一个多项式。

尽管多项式在数学表示式中只占很小的比例，但是它们却非常重要，特别是在代数中更加重要。数学家使用的形容词“代数的”，通常还可以说成“关于多项式的”。看一下代数中的某个定理，甚至是非常高层次上的定理，通过层层分析其意义，很可能就会发现多项式。可以肯定地说，多项式是代数中最重要的概念，无论是在古代还是在现代都是如此。

注　解

[1] 在现代用法中，ℕ 通常包括0。从哲学上我同意这种使用方法。如果你派我到隔壁房间，数一下那个房间里的人数，然后再向你汇报答案，那么“0”是可能的答案。因此0应该包含在这些用于计数的数值当中。然而，因为本书是从历史的角度进行论述，所以我将从 ℕ 中剔除0。

[2] 首先由欧几里得给出了一般的证明，他用的是反证法。假设这件事不为真，即存在某个有理数p/q（其中p和q都是整数），它有这样的性质：$p/q \times p/q=2$。假设p/q是最简分数形式（约去分子分母的公因子，这总能做到），那么p和q中必定有一个是奇数。因为用q乘以上面的等式两边两次得到$p^2=2q^2$，而且只有偶数有偶数平方根，所以p一定是偶数，q一定是奇数。因此对某个整数k，有p是$2k$。于是有$p^2=4k^2$，所以$4k^2=2q^2$，$q^2=2k^2$，因此q也一定是偶数。那么p和q就都是偶数，出现矛盾。因此假设不成立，不存在平方等于2的有理数。

[3] 毕达哥拉斯定理考虑的是一个平面直角三角形的边长。通过简单的观察就可以知道直角的对边一定比其他两个边长。这个定理说的是它的长度的平方等于两个直角边的长度的平方和：$c^2=a^2+b^2$，其中a和b是直角边的长度，而c是对边的长度。这一公式的另外一种表示法如图NP-4所示，是$c=\sqrt{a^2+b^2}$。

第一部分

未知量

第1章

四千年前

按我在引言中所做的粗略定义，作为一种从陈述到质问式算术的思维转变，代数很早就出现在有记载的历史中。我们所知的包含数学的最古老书面文献实际上都含有可以称之为代数的内容。这些文献可以追溯到公元前第二个千年的前半叶，即距现在37或38个世纪之前[1]，它们是由生活在美索不达米亚和埃及的人们写成的。

对于现代的我们，那个世界似乎是无法想象地遥远。公元前1800年距恺撒大帝时代就一如恺撒大帝距我们这个时代那样遥远。在少数专家圈子外，普通人对那个时代、那个区域的了解都来自于《创世记》中所给出的不完整且令人疑窦丛生的记载。对此，受到良好训导的广大西方一神论宗教的信徒们非常了解。这是亚伯拉罕和以撒、雅各和约瑟、吾珥和哈兰、所多玛和蛾摩拉城的世界。那个时代的西方文明包括整个新月沃地地区（见图1-1），它是一大片肥沃的土地，从波斯湾绵延向西北，上行到底格里斯河和幼发拉底河平原，穿越叙利亚高原，然后下行经过巴勒斯坦到尼罗河三角洲和埃及。这个区域的人们彼此相知。整个新月沃地常年有交通往来，从位于幼发拉底河下游的吾珥到位于尼罗河中游的底比斯之间都有交通往来。亚伯拉罕从吾珥到巴勒斯坦然后再到埃及的艰苦旅行，也许正是沿着某些已知的交通路线进行的。

从政治上看，新月沃地的三个主要区域有着非常大的差异。巴勒斯坦是一个穷乡僻壤，是通往他处的交通要道。当时人们把它归属于埃及的

势力范围。埃及是一个种族统一的地区，而且在其边界没有对它形成严重威胁的部落。这个国家在它遭受我后面将讲述的第一次外来侵略之前已经有了1500年的历史，比现在的英格兰还要历史悠久。在自视安全的情况下，埃及人很早就采用了类似一种中国古代封建统治的思维方式，建立了中央集权的君主体制，统治着通过层层选拔人才建立起的庞大的官僚政治机构。早在大约公元前2500年到公元前2350年的第五王朝，就已使用了2000多个头衔。正如罗伯特·威森在《帝国体制》（*The Imperial Order*）一书中所说的那样，“在这种奇妙的等级制度下，每一个人都与其他人不同。”

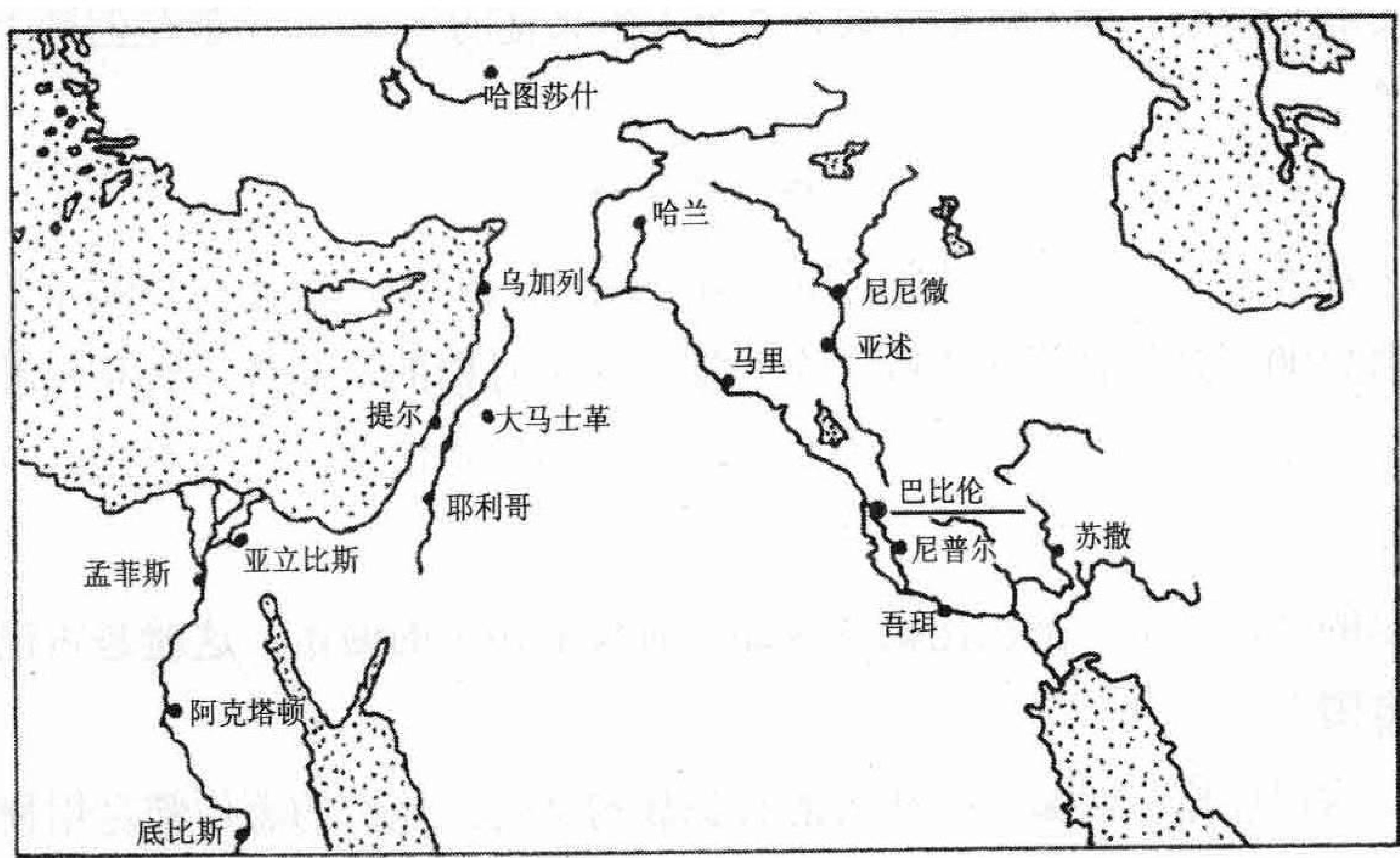
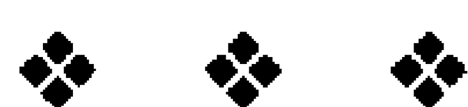


图1-1　新月沃地

而美索不达米亚却呈现出完全不同的景象。这里有太多的民族纷争，首先是苏美尔人，随后依次是阿卡德人、埃兰人、亚摩利人、希泰人、喀西特人、亚述人以及阿拉姆人，他们先后占据统治地位。埃及式的官僚体制也曾经在美索不达米亚占据一时的主导地位，一个强大的统治者可能掌控大面积领土。但是，这些帝国的统治都没有持续很长时间。其中第一个也是最重要的王朝是萨尔贡的大阿卡德王朝，这个王朝统治整个美索不达米亚达160年，大约从公元前2340年到公元前2180年，最后因

高加索部落的袭击而分裂。我要描述的那个时期大约是公元前18世纪到公元前17世纪，此时的萨尔贡的荣耀已经成为褪色的记忆。然而，它给这个地区留下了一种相对通用的语言：闪语族中的阿卡德语。苏美尔语一直是南方的语言，似乎被认为是一种高贵的语言，颇像罗马人说的希腊语或中世纪及近代欧洲说的拉丁语。

然而，美索不达米亚基本上是处在语言和文化上百家争鸣的状态，而不是中央集权控制。这样的环境下创造力到达鼎盛，堪与黄金时代的希腊城邦制、文艺复兴时期的意大利或19世纪的欧洲相媲美。统一都是暂时和短命的。这个时期无疑是“令人感兴趣的”。也许这就是创造力的代价。



在美索不达米亚由不同帝国统治的各个时期中，公元前1790年到公元前1600年是最具影响力的一个时期。这个时期的最高统治者是汉谟拉比，他于这个时期之初在幼发拉底河中部的城邦国家巴比伦登基。汉谟拉比[2]是亚摩利人，说一口阿卡德方言。他把整个美索不达米亚都纳入自己的统治之下，把巴比伦变成那个时代最伟大的城市。这就是古巴比伦帝国[3]。

古巴比伦帝国是一个伟大的有记载的文明。他们的著作都是用楔形文字或称V字形文字的方式写成的。也就是说，写出的字是把成楔形的笔压入到湿粘土而形成的图案。为了长久保存，需要烘烤这些被压过的泥板和圆柱面。很久之前苏美尔人就发明了楔形文字，在萨贡时代引入阿卡德。到了汉谟拉比时代，这种书写方法已经发展成为有600多个符号的书写体系，每一个符号代表一个阿卡德音节。

图1-2是取自于汉谟拉比法典导言中的一个用阿卡德楔形文字写出的短语。这部法典是汉谟拉比在他的帝国强制执行的伟大法律体系。

这个短语的发音类似于En-lil be-el sa-me-e u er-sce-tim，意思是“安利尔，宇宙和地球的统治者”。我们可以从单词be-el看到阿卡德语源自苏

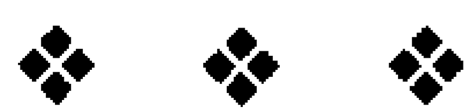
美尔语的事实，它与英语Beelzebub的开头有关，使我们想起了希伯来语的Ba'al Zebhubh，意思是“蝇王”。



图1-2 楔形文字的书写

楔形文字在古巴比伦帝国消亡后仍然沿用了很长时间，直到公元前2世纪。古代的多种语言使用的都是楔形文字。在伊朗的某些古迹上有一些楔形文字的题字，是公元前500年左右居鲁士大帝时代遗留下来的。早在15世纪，这些题字就引起了当时的欧洲游客的注意。从18世纪后期开始，欧洲学者[4]开始尝试着破解这些题字。到了19世纪40年代，人们已经基本掌握了解读楔形文字题字的方法。

大约就在同一时期，很多考古学家开始挖掘美索不达米亚的古迹，例如法国的博塔和英国的奥斯汀·亨利·雷亚德等人。他们发现了大量经过烘烤的刻有楔形文字的泥板。这种考古工作一直持续到今天，现在全世界各地私人收藏或国家收藏总计大约55万块这样的泥板，它们所属的时代是大约公元前3350年到公元前1世纪。这些泥板大都属于汉谟拉比时代，因此形容词“巴比伦的”很随意地就被用在与楔形文字有关的任何事情上，尽管使用楔形文字的历史长达3000多年，而古巴比伦帝国统治时期还不足200年。



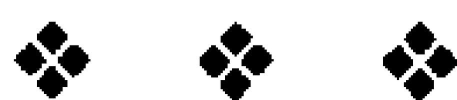
人们很早（至少在19世纪60年代）就已经知道，一些楔形文字泥板中包含数值信息。最早被破译的这类信息都源自人们很容易想到的具有活跃商业传统的组织有序的行政部门，如货物清单、账目等诸如此类的东西。此外还有大量的历法资料。巴比伦人掌握了深奥的历法和广博的天文学知识。

到了20世纪初，出现了很多明显与数学内容有关的泥板，这些内容既不是计时，也与记账无关，让人无法理解。直到1929年奥托·诺伊格

鲍尔开始注意它们时，人们才开始了相关研究。

诺伊格鲍尔是奥地利人，出生于1899年。他参加过第一次世界大战，结果与同乡人路德维希·维特根斯坦①一同被抓入意大利战俘营。战后，他先是成为了一名物理学家，后来转向数学，进入哥廷根大学，在20世纪初最伟大的数学三巨头理查德·柯朗、艾德蒙·朗道和爱米·诺特的指导下学习。到了20世纪20年代中期，诺伊格鲍尔的兴趣开始转向古代数学。他对古埃及做了深入的研究，并发表了一篇关于莱因德纸草书的论文。（稍后我要详细介绍莱因德纸草书。）随后，他又将关注点转向巴比伦，学习阿卡德语，着手研究汉谟拉比时代的泥板。这一工作的成果就是他大约在1935年到1937年出版的三卷巨作《楔形文字数学课本》（*Mathematische Keilschrift-Texte*，德语里keilschrift的意思就是"楔形文字"），巴比伦数学的巨大财富首次在这一著作中得到展示。

纳粹上台后，诺伊格鲍尔离开了德国。他虽然不是犹太人，但是在政治上是自由主义者。哥廷根大学数学学院清除犹太人后，诺伊格鲍尔被任命为该学院的院长。"他担任这一要职只有一天时间，因为他在院长办公室中激烈争辩，拒绝在所谓的忠心声明书上签字。"康斯坦丝·瑞德在《希尔伯特》一书中记述了这一事实。诺伊格鲍尔首先去了丹麦，然后去了美国，在美国他接触到了楔形文字泥板的新藏品。他同美国的亚述专家亚伯拉罕·萨克斯合作，于1945年出版了《楔形文字数学文献》（*Mathematical Cuneiform Texts*）。这本著作现今仍是关于巴比伦数学的英语版权威著作。当然，这方面的研究仍在持续，现在，巴比伦人的辉煌成就已经昭然于天下。而且现在我们知道，他们发明了一些完全可以称得上是代数的技巧。



① 路德维希·维特根斯坦，20世纪最有影响的哲学家之一，语言哲学的奠基人。师从罗素，代表作《哲学研学》。——编者注

诺伊格鲍尔发现，汉谟拉比时代的数学文献有两种：表格文献和问题文献。表格文献就是乘法表、平方表和立方表等的列表，以及一些更高级的列表，例如现存于哥伦比亚大学的“普林顿322”泥板就列出了毕达哥拉斯三元组（即满足$a^2+b^2=c^2$的三元组，根据毕达哥拉斯定理，这三个变量对应于直角三角形的三条边）。

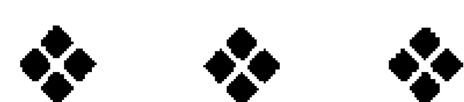
巴比伦人急需这样的表格，因为尽管他们书写数值的系统在当时相当先进，但是却不能像我们熟悉的10个数字那样很方便地做算术。这个体系的进制是60而不是10。例如，我们的数值37表示三个十加上七个一，而巴比伦人的数值37表示三个六十和七个一，相当于我们的数值187。因为缺少“占位的”0，书写数值就更困难。在我们的表示法中有这一符号，所以可以区分284、2 804、208 004等。

小数是如我们的小时、分钟和秒那样书写的，它们本质上是巴比伦人的原创。例如，数值2.5，在这种体系下就被写成2:30。巴比伦人知道在这种体系下，2的平方根大约是1:24:51:10。这个数值就是$1+[24+(51+10\div60)\div60]\div60$，它精确到小数点后6位。与整数一样，占位0的缺失会带来歧义。

即使在表格文献中，代数计算的思想也很明显。例如，我们知道，平方表可用于辅助乘法计算。公式

$$ab=\frac{(a+b)^2-(a-b)^2}{4}$$

把乘法化简为减法（和一个简单的除法）。巴比伦人知道这个公式，或者说他们其实是知道的，只是不知道怎么用这里的方法表示抽象的公式。他们把这个公式看成是一个可以运用于特定数值的步骤，即我们今天所说的算法。



这些表格文献本身非常有趣，但是，只有在问题文献中我们才能看到代数真正的开始。例如，它们包括一元二次方程的解法，甚至特定三

次方程的解法。当然，无论是哪一种解法，都不是写成与现代代数记法类似的形式。所有内容都是涉及实际数值的文字问题。

为了使读者充分领会巴比伦数学，我将用以下三种形式，即实际的楔形文字、文字翻译以及这个问题的现代表示方法，给出《楔形文字数学课本》中的一个问题。

图1-3所示为实际的楔形文字[5]。它被写在一块泥板的两面，我把这两面并排摆放。

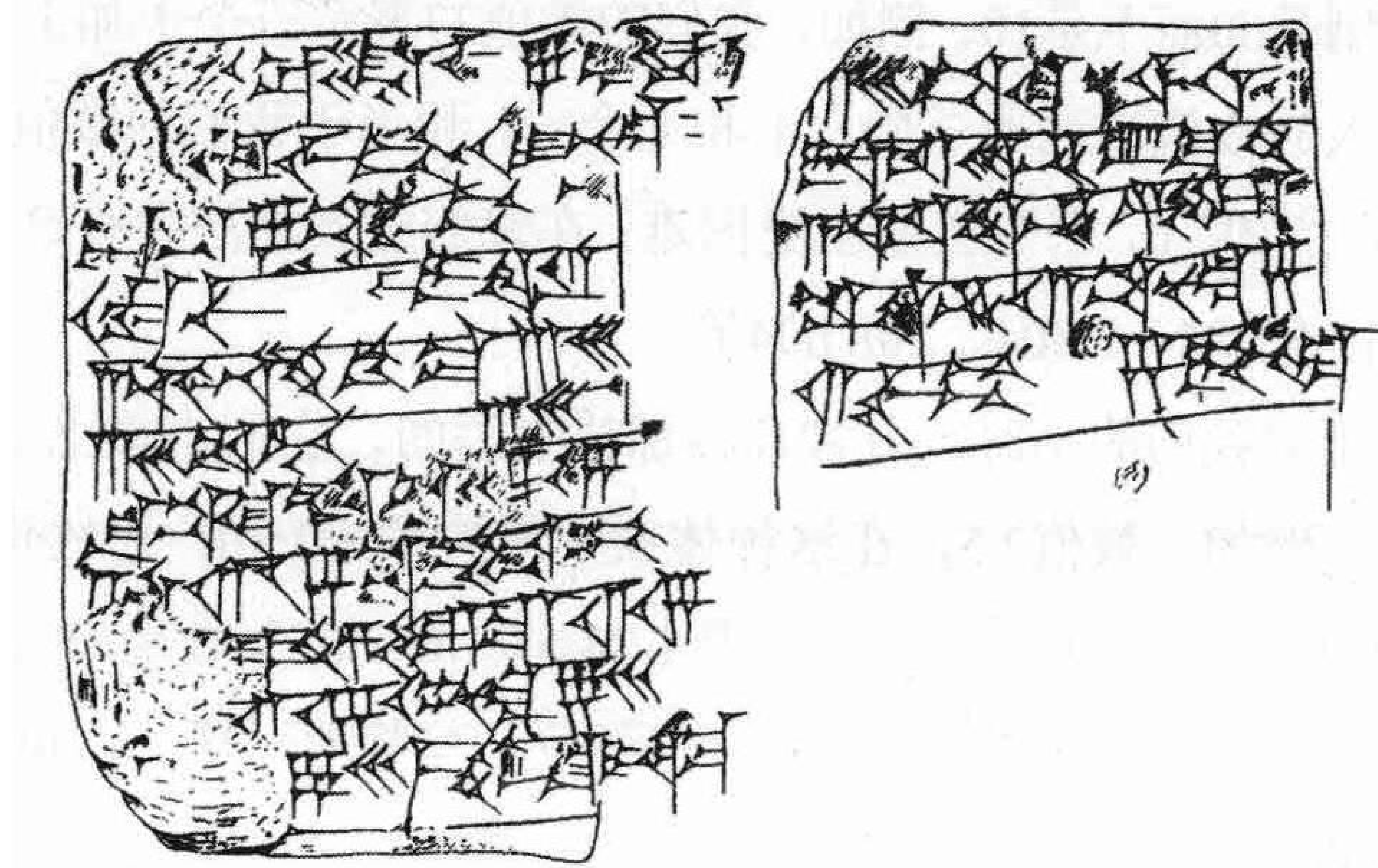


图1-3 楔形文字的问题描述

诺伊格鲍尔和萨克斯把这块泥板的内容翻译如下：（汉字是阿卡德语，字母和数字是苏美尔语，括号里面的部分是不清楚的或根据理解补充的。）

（左图的译文）

[igib]um比igum大7。

[igum和]igibum①是什么呢？

对你来说，二等分7，这是igibum超过igum的值，其结果为3;30。

① igum和igibum是巴比伦数学中表示互为倒数的两个数值的术语。——译者注

把3;30与3;30相乘，其结果是12;15。

对于得到的结果12;15，加上[乘积1,0]，其结果等于1,12;15。

1,12;15的平方根是什么？答案：8;30。

记下[8;30和]8;30，这两个值相等，然后

（右图的译文）

从8;30中减去3;30[得到一个值]

把这两个值相加到另一个值。

一个值是12，另一个值是5。

于是igibum等于12，igum等于5。

（注意：诺伊格鲍尔和萨克斯使用逗号分开这里的数值的数位，使用分号分隔整数部分和小数部分。所以1,12;15表示（1×60+12+15/60），即72/4。）

下面是用现代方法求解这个问题的过程。

一个数比它的倒数大7。（注意，因为巴比伦数的位-值存在歧义，所以x的“倒数”可能是$1/x$、$60/x$、$3600/x$等。事实上，可能是60的任何次幂除以x。但从文献作者的求解过程可以看出，在这里作者所取的“倒数”应该是$60/x$。）而

$$x-\frac{60}{x}=7$$

x和它的“倒数”是什么呢？因为上面这个方程可以简化成

$$x^2-7x-60=0$$

我们能够运用熟悉的公式[6]得到

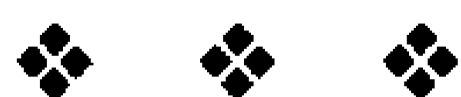
$$x=\frac{7\pm\sqrt{7^2+(4\times 60)}}{2}$$

得出x=12或x=−5。巴比伦人不知道负数，3000年后人们才开始普遍使用负数。所以他们关心的解只有12，它的“倒数”（即$60/x$）

是5。事实上，巴比伦人的算法不能得出上面二次方程的两个解，而是等价于求x和它的“倒数”的一个略微不同的公式

$$x=\sqrt{\left(\frac{7}{2}\right)^2+60}\pm\frac{7}{2}$$

如果你对此吹毛求疵，可以说，这表明严格地说他们没有解决这个二次方程。尽管如此，你也不得不承认这是青铜器时代早期数学的一个相当漂亮且意义重大的成果。



我再强调一次，汉谟拉比时代的巴比伦人没有真正的代数符号体系。他们所用的是文字，使用一种原始编号系统来表示的量。在使用“未知量”进行思考方面他们仅仅是向前迈进了一些，在阿卡德语文献中用苏美尔语的单词表示未知量，例如上面的问题中的igum和igibum。（诺伊格鲍尔和萨克斯把igum和igibum都翻译成“倒数”。在其他一些文献中则使用了苏美尔语来表示矩形的“长”和“宽”。）这些算法不是普遍适用的，对不同的文字问题使用不同的算法。

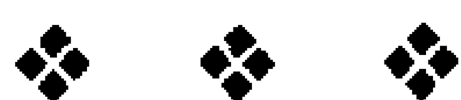
由此产生两个问题：第一，他们为什么这样做？第二，是谁首先解决了这些问题？

对于第一个问题，巴比伦人当然没有打算告诉后人他们为什么这样做。最靠谱的猜测是，这些文字问题可能是作为一种检查计算的方法而出现的，计算可能包括土地面积的测量，而问题可能是挖一定尺度的河渠时被挖出的泥土量。当人们圈出一块长方形土地并计算它的面积时，他们可以“反过来”通过某个二次方程的算法求它的面积和周长来确认得到的是正确的结果。

对于第二个问题，汉谟拉比时代泥板中出现的原始代数是非常成熟的。根据我们了解的远古时代知识进步的速度可知，这些技术也许酝酿了好几个世纪。是谁首先想出来的呢？这个我们无从知道，然而在这些问题表格中对苏美尔语的使用表明苏美尔人是其创始者。（同现代数学

中希腊字母的使用一样。）我们有汉谟拉比时代之前的文献，也就是3000年之前的资料，但它们都是算术文献。也只有在公元前18世纪和公元前17世纪的时候，代数思想才开始出现。假定存在可以显示这些代数思想的更早成型的“过渡”文献，那么或者它们没有保存下来，或者至今尚未找到。

我们在汉谟拉比时代的泥板上也没有发现写出这些原始代数的人的任何蛛丝马迹。我们知道大量巴比伦人的数学成果，却连一名巴比伦数学家也不知道。我们知道其名字的第一个人生活在新月沃地的另一端，他极有可能是一位数学家。



正当汉谟拉比在美索不达米亚的统治日益巩固的时候，埃及正在遭受它的第一次外来入侵。这些侵略者在希腊语中被称为*西克索斯*，这个单词是埃及语的短语“外来统治者”的讹误。这些外来者是从巴勒斯坦开始入侵的，他们没有采取突然袭击，而是慢慢地吞并并将其殖民化。大约在公元前1720年，他们在尼罗河三角洲东部的阿瓦利斯建都。

在西克索斯王朝，有一位名叫阿梅斯的人，他就是我们知道名字的第一位与数学有着实在联系的人物。至于他是否是职业数学家尚无法确定。我们是通过一份纸草书知道他的，这份纸草书大约是公元前1650年的产物，那还是西克索斯王朝的初期。这份纸草书表明阿梅斯是一名抄写员，抄写在第十二王朝（大约公元前1990—前1780）写下的一份文书。这是我们知道的源于西克索斯王朝的文献之一，西克索斯的统治者们非常崇拜当时的古埃及文明。也许阿梅斯是一位数学外行，只是盲目地抄写他看到的文字。然而，这似乎不太可能。这份纸草书中有几处错误，而这些错误看起来更像是计算错误（后面的计算用的一直是错误的数值），而不是抄写错误。

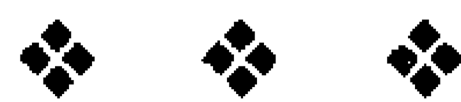
这一文书通常被称为莱因德纸草书，以纪念苏格兰人亨利·莱因德。莱因德患有肺结核，于1858年到埃及休假，期间在卢克索城买了一份纸

草书。在他死后五年，大不列颠博物馆得到了这份纸草书。现在，人们认为应该以这份纸草书的作者的名字为其命名，而不是以买了它的人的名字为其命名，因此现在通常也称其为阿梅斯纸草书。

虽然这是数学史上令人兴奋的伟大发现，但在阿梅斯纸草书中仅包含几个我所讨论的意义下的代数思想的明显线索。下面是这份纸草书中的问题24——代数问题："一个量加上它自身的四分之一等于15。"我们用现代记法写出来，就是：

$$x+\frac{1}{4}x=15$$

并求未知数x。阿梅斯运用了试错法。这份纸草书中的巴比伦风格的系统化算法寥寥无几。



詹姆士·R. 纽曼在《数学世界》(*The World of Mathematics*)中写道："关于埃及的数学发展程度，在学习古代科学的学生们之间存在明显不同的观点。"这种观点上的不同现在依然存在。在研读了巴比伦和埃及的代表性文献之后，我不明白为什么还有人主张公元前1750年左右分别在新月沃地两端繁荣起来的这两个文明古国在数学发展水平上是相当的。尽管二者都以算术的风格进行研究，而且也没有证据显示他们拥有抽象的能力，但是显然巴比伦问题在深度与精确度上都要领先于埃及问题。(顺便说一下，这也是诺伊格鲍尔的观点。)

这些古代的人们仅使用最原始的数值书写方法就取得了如此辉煌的成就，真是一件了不起的事情。更令人惊叹的是，他们的成就甚至随后的几个世纪都没有人超越。

注　解

[1] 美索不达米亚早期历史的时间至今没有确定。在本书撰写时期，经常引用"中间"年表，这也是我所用的年表。还有低年表、超低年表和高年表。发生在"中间"年表中公元前2000年的事件在高年表中的时间是公元前2056年，在低

年表中是公元前1936年，在超低年表中则是公元前1911年。无疑，亚述专家为争论这些不惜同好友断绝往来，甚至夫妻感情破裂。我倒没有什么意见，精确的年代对我的研究不重要。总之，1950年之前所写的大部分资料中的日期推算都不太可信。

[2] 汉谟拉比（Hammurabi）的另一个常用拼写是Hammurapi。古英语文献使用的是Khammurabi、Ammurapi和Khammuram。然而，认为汉谟拉比是《创世记》中的Amraphel的观点现在已被否认。亚伯拉罕的时代都是推测的，但是人们似乎都认为他生活的年代要比与汉谟拉比统治的时代早。

[3] 西方历史更熟悉的是第二帝国。第二巴比伦帝国指的就是尼布甲尼撒二世的帝国，当时犹太人被监禁。丹尼尔服务的是同一位君主。描述伯沙撒森林战争的作品暗示第二巴比伦帝国败给了波斯人。这一切发生在汉谟拉比时代的一千年后，本文不予介绍。

[4] 其中几个重要人物是丹麦人卡尔斯顿・尼布尔、德国人乔治・弗里德里・希格罗特芬德和英国人亨利・罗林森爵士。顺便提一下，希格罗特芬德应邀参加了位于哥廷根的汉诺威大学组织的楔形文字的破译工作，后来这所大学成为闻名于世界的优秀数学中心。

[5] 事实上楔形文字也不都那么难理解。了解楔形文字编号的最佳参考书是约翰・康威和理查德・盖合著的《数值之书》。

[6] 还是给不熟悉二次方程求根公式的读者介绍一下吧。二次方程$x^2+px+q=0$有两个解：

$$x=\frac{-p\pm\sqrt{p^2-4q}}{2}$$

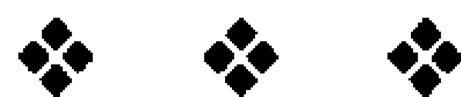
第2章

代数之父

接着说埃及。代数之父丢番图[1]生活在1世纪、2世纪或3世纪的罗马统治下的埃及的亚历山大城，我以他的这一荣誉称号作为本章的标题。

丢番图是不是代数之父正是律师们所谓的“难以决断的问题”。若干非常令人尊敬的数学史学家都否认这一点。例如，在科学传记辞典中，库尔特·沃格尔认为丢番图的工作并不比古巴比伦人和阿基米德（公元前3世纪，参见下文）的工作更代数化，因此得出结论说：“丢番图肯定不能像人们通常所称呼他的那样是代数之父。”范德瓦尔登把代数的诞生时间往后推迟了一段，认为数学家花拉子米才是代数之父。花拉子米比丢番图晚出生600年，我将在下一章介绍他。另外，现在本科生所学的丢番图分析这一数学分支通常是作为数论课程的一部分，而不是在代数里讲述。

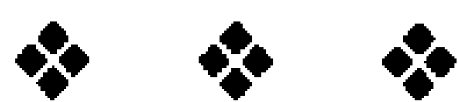
我将讲述丢番图的工作，并对此提出我自己的观点，然后读者可以根据他的工作价值给出自己的判断。



在公元前141年整个地区被帕提亚人征服之前，美索不达米亚人在持续了几个世纪的种族和政治纷争中一直使用楔形文字进行书写。我们现在还保留有这次征服之前用楔形文字书写的数学文献。汉谟拉比帝国从帕提亚人的征服中挣脱出来之后的1500年间，在数学符号体系、技术和认知方面几乎没有任何进步，这已得到研究过这一课题的每一个人的考

证，是一件令人非常惊讶的事情。对楔形文字做过研究的数学家约翰·康威说，从泥板自身看得出来的唯一差异就是最新的泥板中有“占位零”的标记，即一种可以区分如281和2801的方法。埃及同美索不达米亚一样：我们没有证据可以认为，埃及数学从公元前16世纪到公元前4世纪有什么令人瞩目的进步。

即使巴比伦和埃及的数学家在自己的国土上没有取得什么进步，但是他们早期的辉煌发现已传遍古代西方，而且可能传得更远。而从公元前6世纪开始，古代世界中代数的历史主要是一段希腊的历史。



希腊数学的独特之处在于，在丢番图之前，它主要研究的是几何。对此给出的理由在我看来似乎有些道理，这一理由是毕达哥拉斯学派（公元前6世纪后期）有在数值的基础上建立数学、音乐和天文学的思想，但是无理数的发现困扰了这些毕达哥拉斯人，所以他们将兴趣从算术转向了几何，因为算术似乎包含一些无法书写的数值，而在几何中这样的数值可以准确无误地用线段长度表示。

因此，早期希腊代数概念都是用几何形式表示的，通常非常晦涩难懂。例如，巴什马科娃和斯米诺娃指出，欧几里得的伟大著作《几何原本》第6卷中的命题28和命题29给出了二次方程的求解方法。我认为它们确实给出了二次方程的解决方法，但至少在第一次看它们时，这不容易发现。下面是托马斯·希思爵士翻译的欧几里得命题28[①]。

在已知线段上作一个等于已知直线图形的平行四边形，它

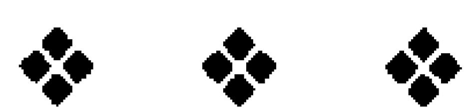
① 这一命题的描述很晦涩，根据陕西科学技术出版社出版的《几何原理》中译本中该命题及其证明的描述，用现代语言描述该命题如下。给定线段AB，直线图形C（可以理解为任意多边形）和平行四边形D。设AB的中点为E，且当作相似于D的平行四边形$EBFG$时，C的面积不大于$EBFG$的面积。这时，在线段AB（的一个部分）上（可）做一个平行四边形$AHIJ$（HI是平行四边形的一条边），它的面积等于C的面积，而且由H、I和B决定的平行四边形相似于D。——译者注

> 是由取掉了相似于某个已知平行四边形的平行四边形而成的：这个已知直线图形必须不大于在原线段一半上相似于取掉的图形的平行四边形。

明白了吗？巴什马科娃和斯米诺娃说这等价于求解二次方程$x(a-x)=S$①。我同意他们对此问题的理解。

欧几里得在亚历山大将军托勒密（他的王朝是托勒密I世，统治时期是公元前306年到公元前283年）统治下的亚历山大城生活。当时亚历山大城是埃及的一部分，欧几里得在那里建立了学校，开始传道解惑。在欧几里得出生前不久，亚历山大在尼罗河三角洲的西岸建立了亚历山大城，它隔着地中海与希腊遥遥相望。人们认为欧几里得在埃及定居之前已经在雅典的柏拉图学园接受过数学训练。不管怎样，公元前3世纪的亚历山大城是伟大、发达的数学中心，比希腊本身更加重要。

阿基米德比欧几里得年轻40岁，他可能在亚历山大城欧几里得的后辈的指导下学习，学习的仍然是几何方法，尽管他把这一方法用于更加不易处理的领域中。例如，他的著作《论劈锥曲面体和球体》探讨的就是平面与一种经典的弯曲二维表面的交集。从这一著作可以清楚知道，阿基米德能够求解特定种类的三次方程，就如同欧几里得能够求解一些二次方程一样，但是其使用的语言都是几何语言。



经过公元前3世纪的辉煌后，亚历山大学派的数学发展出现了某种程度的下滑。到了混乱的公元前1世纪（想想安东尼和埃及艳后克莉奥帕特拉便知），亚历山大学派的数学几乎完全消失。后来，由于早期罗马帝国更加稳定等诸多条件，数学有了一定程度的复苏。同时也出现了脱离纯几何的思维转变，丢番图就生活、工作在这样一个新时代。

---

① 对于上面注释，当AB的长度为a，D的长度为1的正方形，且C的面积为S时，命题28相当于求HB的长度x，使得矩形$AHIJ$的面积为$(a-x)x=S$。——译者注

正如我在本章开始时所声明的那样，对于丢番图我们几乎一无所知，甚至不知道他生活在哪个世纪。人们最喜欢推测的世纪就是公元3世纪，其中引证的时间大多是公元200年到284年。我们是因为到丢番图写了一篇名为《算术》的论文而注意到他，该论文只有一大半遗留了下来。这篇论文保留下的部分由189个问题组成，其主题是寻找满足一定条件的数值或一类数值的问题。这篇论文的开头是一个介绍，丢番图概括介绍了他的符号体系和方法。

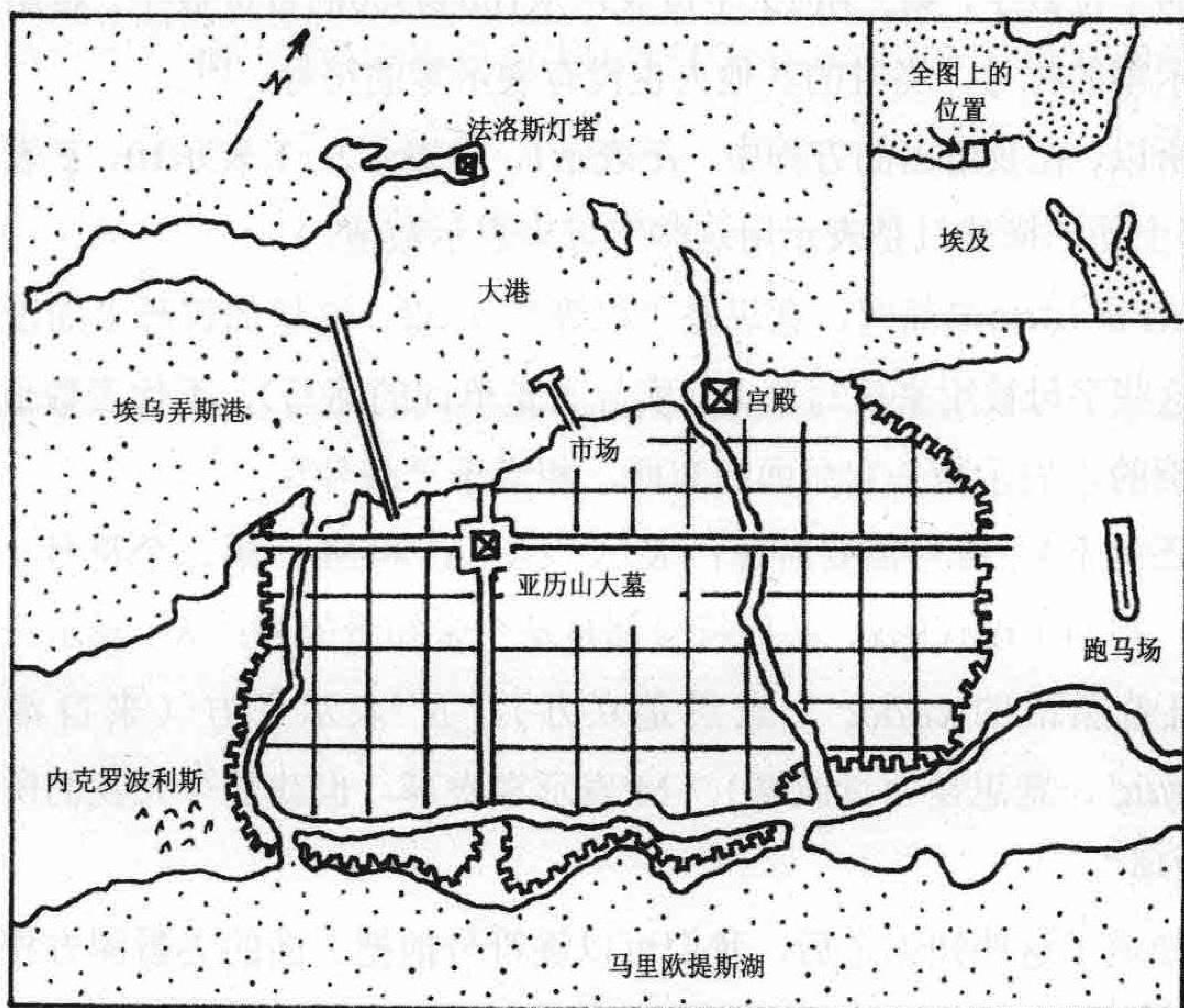


图2-1 古亚历山大城。法洛斯灯塔曾是著名的灯塔，它是古代世界七大奇迹之一，毁于从7世纪到14世纪的一系列地震之中。人们认为那个伟大的图书馆就在这个城市东北方的宫殿的附近。锯齿形线展示了原来的城墙（公元前331年）

在我们看来，他的符号体系相当简陋，但是就其所处的时代来说，这已经很了不起了。我们用示例介绍其重要的几点。下面是现代形式的一个方程：

$$x^3 - 2x^2 + 10x - 1 = 5$$

丢番图把它写成如下形式：

$$K^{Y}\overline{\alpha}\varsigma\overline{\iota} \pitchfork \Delta^{Y}\overline{\beta}\,\mathrm{M}\,\overline{\alpha}\,'\acute{\iota}\sigma\,\mathrm{M}\overline{\varepsilon}$$

最容易挑出的东西就是数值。丢番图使用希腊“字母”体系来书写数值。其作法是采用希腊字母表中24个通用字母，然后再增加3个已废弃的字母，一共27个符号。把这些字母每9个一组分成3组。这扩充了的字母表的第一组9个字母表示从1到9的个位数字，第二组9个字母表示从10到90的十位数字，第三组9个字母表示从100到900的百位数字。希腊人没有表示零的符号，当时的其他人也没有表示零的符号。[2]

所以，在我给出的方程中，$\overline{\alpha}$表示1，$\overline{\beta}$表示2，$\overline{\iota}$表示10，$\overline{\varepsilon}$表示5。（字母上面的横线只是表示用这些字母来表示数值。）

$'\acute{\iota}\sigma$是$'\acute{\iota}\sigma o\varsigma$的简写，意思是“相等”。注意，这里的符号上面没有横线，这些字母被用来拼写单词（实际上是单词的缩写），不代表数值。这个颠倒的$\pitchfork$表示减去它后面的东西，相当于“减号”。

还剩下4个符号需要解释：K^{Y}、ς、Δ^{Y}和$\dot{\mathrm{M}}$。第二个符号ς是未知量，相当于现代的x。其他符号都是这个未知量的幂：K^{Y}表示三次幂（来自希腊语的$\kappa\acute{\upsilon}\beta o\varsigma$，意思是立方），$\Delta^{Y}$表示平方（来自希腊语$\delta\acute{\upsilon}\nu\alpha\mu\iota\varsigma$，意思是力量或幂），$\dot{\mathrm{M}}$表示零次幂，也就是今天我们所说的“常数项”。

知道了这些知识之后，我们可以逐符号地把上面的丢番图方程翻译如下：

$$x^3 1 x 10 - x^2 2 x^0 1 = x^0 5$$

如果加上隐藏的加号和某些括号，这个方程的意思就更加清晰了：

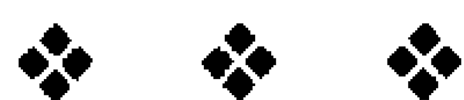
$$\left(x^3 1 + x 10\right) - \left(x^2 2 + x^0 1\right) = x^0 5$$

因为丢番图把系数写在变量的后面，而不是像我们那样把系数写在前面（我们写$10x$，而他写成$x10$），而且因为任何数的零次幂是1，所以上面这个方程等价于我最先书写的那个样子：

$$x^3-2x^2+10x-1=5$$

从这个例子可以看到，丢番图在进行处理时有相当娴熟的代数记法。我们只是不清楚这样的记法有多少是他原创的。使用特殊符号表示未知量的平方和立方可能是丢番图自己发明的。然而用ς表示未知量的使用方法[3]似乎是从一位更早的作者那儿学来的，这位作者是密歇根大学收藏的密歇根纸草书620的作者。

丢番图的记法体系也有一些缺点。一个主要的缺点是不能表示第二个未知量。用现在的话说，有x但是没有y或者z。这是丢番图面临的一个主要困难，因为，虽然不像高斯误以为的那样他著作研究的都是不定方程，但不定方程仍是主要部分。这需要稍微解释一下。



数学家使用“方程”一词表示这样一种意思：某种东西等于另外一种东西。如果我说“二加二等于四”，就已经描述了一个方程。当然，包括丢番图在内的数学家感兴趣的是与其中某些未知量有关的方程。未知量的存在把一个方程从陈述语气“是如此”变成疑问语气“是如此吗”，或者更常见的“什么时候如此”。下面的方程

$$x+2=4$$

隐含着这样一个问题：“什么加上二等于四？”答案当然是2。所以这个方程在$x=2$时成立。

然而，假设我问下面的方程的解是什么，答案就没那么明显了。

$$x+y+2=4$$

乍看起来，数学家立即想要知道你要寻找什么样的答案。仅是正整数吗？那样的话，唯一的答案是$x=1$，$y=1$。你接受非负整数吗（即包含0）？如果接受，那么我们有两个解：$x=0$，$y=2$；$x=2$，$y=0$。你允许解是负数吗？如果你允许，那么就有无数多个解，例如$x=999$，$y=-997$。你允许解是有理数吗？如果你允许，那么又有无数多个解，例如$x=157/111$，$y=65/111$。当然，如果你允许解是无理数或者复数，那么解的数量就会更多了。

诸如此类的方程中有多个未知量，而且解的数目有无穷多个（解的多少取决于要求的解的类型），这样的方程称为*不定方程*。

最著名的不定方程就是费马大定理中的不定方程：

$$x^n + y^n = z^n$$

上面方程中的x、y、z都是正整数。如果n=1或者n=2，那么这个方程有无穷多个解。费马大定理说当n大于2时，这个方程没有正整数解。

费马大约在1637年阅读了丢番图的《算术》（拉丁语的翻译本），当时他就想到了这个定理，于是他在这本书的空白处留下了那条著名的注释，陈述了这一定理，然后又加了下面一句话（也是用拉丁语）："对此我已经找到一个完美的证明，但是这一空白地太小，写不下。"这个定理直到357年后才得到了证明，是由安德鲁·威尔斯证明的。



可见，《算术》一书大部分内容讨论的都是不定方程。而且，这使丢番图处于非常不利的境地，因为他只有一个表示未知量的符号（未知量的平方、立方等有额外的符号）。

为了弄明白他是如何克服这一困难的，我们来看看他的求解方法——费马就是在这个问题的空白处写下了他的著名笔记：卷2的问题8。

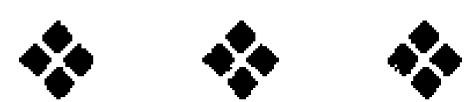
丢番图叙述了这样一个问题："把一个平方数分解成平方数的和。"今天，我们会把这一问题表示成如下形式："给定一个数a，寻找数x和y，使得$x^2+y^2=a^2$。"当然，丢番图没有我们这样的巧妙记法，所以他只能用语言描述这一问题。

为了解决这一问题，他开始时给a一个定值4。所以我们开始寻找符合$x^2+y^2=16$的x和y。然后他用x写出一个看起来随意的特殊表达式来表示y，$y=2x-4$。于是，现在我们要解一个特殊的方程，一个对丢番图来说可以使用他的字符体系的方程：

$$x^2 + (2x-4)^2 = 16$$

这正是一个二次方程，丢番图知道如何求解它。这个解是x=16/5。（还有

另一个解：x=0。因为丢番图没有表示零的符号，所以他舍弃了这个解。）于是y=12/5。这个解看起来似乎不会令人有什么深刻的印象。事实上，它有点骗人的嫌疑。方程$x^2+y^2=a^2$有无穷多个解。丢番图只得到一个解，然而他采用了一个可以一般化的方法得到了这个解，正如他在这本书的其他地方所说的那样，他知道这个方程有无穷多个解。



之前我提到过，当数学家看到不定方程$x+y+2=4$时，他们脑海中的第一个问题是：你要寻找什么类型的解？对于丢番图的方程来说，答案是正有理数解，例如刚才计算出来的解16/5和12/5。那时人们还没有发现负数和零，所以丢番图会说方程$4x+20=4$是“荒谬的”。当然，他会知道无理数，但对此不感兴趣。当这些数出现在一个问题中时，他会调整这个问题的条件，以便得到有有理数解的方程。

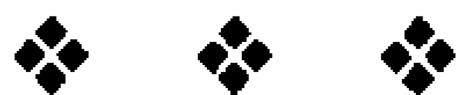
寻找如丢番图处理的问题的有理数解，实际上等价于寻找密切相关问题的整数解。等式

$$\left(\frac{16}{5}\right)^2+\left(\frac{12}{5}\right)^2=4^2$$

可以简化为

$$16^2+12^2=20^2$$

所以，今天我们使用短语“丢番图分析”表示“寻找多项式方程的整数解”。



也许读者对于丢番图对方程$x^2+y^2=a^2$的处理仍旧没有什么感觉，其中几乎没有展示出2000年来巴比伦人对二次方程的解和他们其他成就的非凡魅力。

为了更公平地对待丢番图，我应该说，虽然那个特殊的问题是丢番图提出的一个比较容易的问题，而实际上他所解决的问题要比这道题难得多。方程$x^2+y^2=a^2$非常好地说明了他的方法，这个方法与费马大定理有

着有趣的联系，这不需要过多解释。这就是为什么它是一个受欢迎的例子。除此之外，丢番图还研究了一个未知量的三次或四次方程，两个、三个或四个未知数的联立方程组，以及一个相当于12个未知量的8个方程的联立方程组的问题。

丢番图在解决方程$x^2+y^2=a^2$的过程中展示出他掌握了更多东西。例如，他知道符号法则，他是这样描述的：

> 所需的（wanting）（即负数）乘以所需的等于现成的（forthcoming）（即正数），所需的乘以现成的等于所需的。

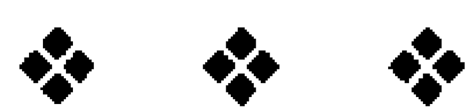
我早就提到过，在负数还没有被发现的情况下知道这一事实，真是一件了不起的事情！

事实上，上文所说的“没有被发现”需要附加一些条件。尽管丢番图没有发现负数是一个独立数学对象的概念（至少他没有负数的任何记法），而且他也不把负数当作解，但是实际上他却在计算中自由地使用了负数，例如，(x^2+4x+1)减去$(2x+7)$等于(x^2+2x-6)。显然，即使他认为-6本身作为数学对象是“荒谬的”，但是在某种层次上他知道$1-7=-6$。

正是这类情况，让我们认识到数学思想是多么怪异。就是负数这样一个基本概念也花了几个世纪才在数学家的头脑中弄清楚，其过程中有很多类似上面所说的中间阶段。此后大约1300年，人们才发明了虚数。

丢番图还知道如何从方程的一边到另一边做“移项”操作，即改变这些项的符号；知道为了化简如何“合并同类项”以及展开和因式分解等一些初等法则。

他对有理数解的执着已经触及了我们现代的代数数论。我们今天会说方程$x^2+y^2=a^2$是半径为a的圆的方程。在寻找这个方程的有理数解的时候，丢番图问：有理数坐标x和y在这个圆的什么地方？正如我们将在第14章所述的那样，这的确是一个非常现代的问题。



因此就说丢番图是代数之父吗？我愿意给他这样的荣誉，正是因为他的文字符号体系——他对未知量、未知量的幂、减法及相等而使用的特殊的符号等。当我第一次看到丢番图完全用他自己的符号写出的一个方程时，我的第一反应，也可能是你的第一反应，是“他说的是什么？”接着再往下看，我很快就熟悉了他的符号体系，甚至能够不加思索地快速阅读丢番图的方程。

最终，我领会到丢番图制造出的符号体系是多么先进。我同意沃格尔的观点——《算术》一书在一般方法上有缺失，我也愿意相信丢番图在选择话题上缺乏原创性。也许他不是第一个对未知量使用特殊符号的人吧。

然而，丢番图最早把如此广泛的各种问题收集起来传给我们，这是历史财富。我们不知道是谁第一个使用符号表示未知量，因此而感到羞愧，但是丢番图这么早就能如此好地使用符号来表示未知量，我们又为他感到荣耀。也许有个我们从来不曾了解的人才是真正的代数之父。但是既然这个头衔空着，我们最好还是把它送给一个最有资格的人，这个人就是丢番图。



图2-2　亚历山大城的法洛斯灯塔，这是马丁·赫姆斯科克（1498—1574）根据想象绘制的

注 解

[1] 丢番图（Diophantus）把他自己的名字写成希腊语的形式Diophantos。欧洲人是通过拉丁语译本才广泛了解他的著作的，因此沿用他的拉丁语名字Diophantus。

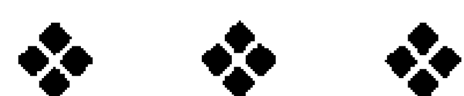
[2] 因此，例如，ψμθ 也许是749。用作单位的字母可以再利用，来表示几千。例如，δψμθ 的意思是4 749。δ 通常的意思是4，在这里它表示4 000。对于超过9 999的数，数字被分成4个一组，用M（代表Myriad，“无数的”）或丢番图的点记法分开。例如，数值δτοβ·ηϡζ应该是43 728 907。（那个看起来有点奇怪的字母ϡ已经废除，在这里用于表示900。因为ζ代表7，所以ϡζ是907。注意这里占位的零，因为这种方法不需要零。）

[3] 在这一使用方法中，希腊字母表中的这个用于词尾的字母ς 的头顶上有一个小叉。这个字母打印不出来。密歇根纸草书要推算到公元2世纪早期，要比普遍认可的丢番图时代大约早一个世纪。

第3章

还原与化简

众所周知，“代数”一词源自阿拉伯语。依我看，这多少有点儿不公平。我将简单讲述其中的原因。到底公平与否，还需要听一听历史学家的解释。



假设我所给出的丢番图的生存年代（公元200—284）是正确的，那么他所处的时代是一个非常不幸的时代。当时的埃及是罗马帝国的一个省，它正处于众所周知的衰落时期，也就是爱德华·吉本花费大量笔墨描述的那一时期。吉本著作[1]的第7章详细地描述了那个糟糕的时期，也就是丢番图生活的时代。

罗马帝国在3世纪后期重震旗鼓。戴克里先大帝（公元284—305）和康斯坦丁大帝（公元306—337）都是最有名的罗马皇帝。前者疯狂地迫害基督教徒；后者是一位基督教徒的儿子，发布了米兰（公元313年）法令，命令整个帝国包容基督教，并在病危时接受了洗礼。

从起初对基督教的宽容到随后的强迫实施基督教的决策，都无法阻止罗马帝国的衰亡。在一定程度上，这也许加速了它的衰亡。早期基督

---

① 指吉本所著的《罗马帝国兴衰史》。吉本是英国著名历史学家，他于1772年开始写作此书。该书共6卷，上下纵横1300年，收尽罗马帝国的变幻风云。西方人每言及罗马，必称此书。——编者注

教最大的优点之一是它面向所有阶层。然而，为了做到这一点，它必须使用复杂而有感染力的形而上学理论“收买”世故的都市知识分子，同时利用简明而直白的救世和神权的信条来维护它对大众的控制，并利用五花八门的传闻或者对根深蒂固的某些异教信仰做些让步来加强控制。不过，大众仍然会通过各种途径察觉到对这些高深的形而上学理论的争论，并利用这些争论掩饰社会或种族不满的情绪。

丢番图所居住的亚历山大城就是这一过程的很好例证。经历了300年的希腊统治，又经历了300年的罗马统治，亚历山大城一直是一块华丽辉煌的海外飞地，不开化并说着科普特语的埃及农夫们居住的穷乡僻壤是它的粮食和衣物供应地。对于来自于沙漠边缘的基督教科普特人来说，“希腊”、“罗马”和“异教徒”也许是近义词，而缪斯女神神殿与其经久不衰的学术传统和令人着迷的伟大图书馆，似乎就是撒旦的家。

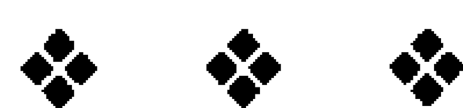
禁欲主义的宗教信仰使得埃及的情况更糟。在亚历山大城，这一宗教崇拜特别强烈，这里安置了几千名身强力壮且性饥渴的男性，供那些想要抽打宗教罪犯的人差遣。野心勃勃的政治家就喜欢干这种勾当。在当时的罗马，政治家中经常有教会的官员。这就是公元415年海帕希亚被谋杀的背景，这也是令爱德华·吉本怒不可遏的事情。

海帕希亚是数学史上第一位女性。所有关于她的文献都已经丢失了，我们只能通过传说了解她的情况，因此我们很难判断她是否是一位重量级的数学家。但有一点可以肯定，她是一位优秀的知识分子。她在缪斯神殿授课（他父亲塞昂是最后一任校长），既是教科书的编写人又是编辑，同时还是保管员。这些教科书中就有数学教科书。她教授新柏拉图派哲学[1]，也是该学派的忠实信徒这一哲学试图确立在后罗马时代非常缺乏的秩序、正义和和平。据说她非常美丽而且是一名处女。

海帕希亚在教学和研究等方面非常活跃，当时西里尔是亚历山大城的大教主。西里尔后来被称为圣西里尔，由于时代久远且神学争论颇复杂，我们很难断定他到底是什么样的人。吉本对他颇多微辞，但吉本对

基督教抱有一定的偏见，因此对他的话我们不能完全相信。西里尔肯定发起了针对亚历山大犹太人的大规模屠杀，把他们驱逐出这座城市，但是犹太人好像在此之前就发起了一次令人不齿的反基督屠杀，甚至连吉本都对此予以承认。[2]正如我们从《天主教百科全书》中了解到的那样，当时的亚历山大城总是处于“暴乱”之中。总之，西里尔卷入了与驻埃及罗马提督奥列斯特之间的一场国家与教会关系的争论之中，有人散布谣言说海帕希亚是修复这一裂痕的主要障碍。一群暴徒被煽动起来，也许是自行发起暴动，他们把海帕希亚从马车里拖出来，一直拖过几条大街来到教堂，据文献记载及其译文描述，她被用贝壳或者陶瓷碎片凌迟处死。[3]

海帕希亚似乎是在缪斯神殿授课的最后一个人，人们认为415年她的死标志着古老的欧洲数学的结束。西方的罗马帝国又勉强支撑了60年，亚历山大城继续在东方拜占庭各皇帝的统治下存活了164年（中间因波斯人[4]的入侵而间断，时间为616～629年），但思想活跃的生命力却一去不复返。代数历史中下一位著名人物的家乡在亚历山大城以东900英里的底格里斯河的沿岸，又回到了它2500年前诞生的美索不达米亚平原。



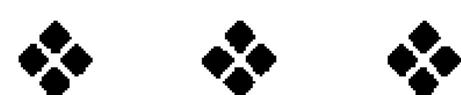
公元5世纪，罗马帝国的北边和西边领土都被日尔曼人攻占。南边和东边的领土除了希腊、安纳托利亚及意大利和巴尔干半岛以南的领土外都在7世纪沦陷于伊斯兰教大军的手里。亚历山大城本身也于公元640年12月23日沦陷，被拜占庭皇帝赫拉克利乌斯[5]攻破，此人穷其毕生的精力试图夺回7世纪初被波斯人侵占的领土。

亚历山大城实际的征服者是一位名叫阿姆鲁・伊本・阿斯[6]的人。他是第二任哈里发（即穆罕默德死后穆斯林的领导人）奥马尔的亲信干将。亚历山大城发起了反攻，围城持续了14个月，而埃及的大部分地区都被穆斯林轻松拿下。当时埃及人一直信奉基督教的所谓基督一性说[7]，暂时摆脱了波斯人的奴役之后，拜占庭皇帝赫拉克利乌斯却因此对他们

施加残酷的迫害。其结果是埃及本土人开始憎恨拜占庭人，非常高兴有一个更加宽容的统治者来取代目前这个残酷的统治者。

第三任哈里发奥斯曼属于穆罕默德部落的伍麦耶王朝分支。在与先知的女婿阿里之间经过多次混乱的内战之后（一个间接结果是逊尼派教徒和什叶派教徒的分裂，直到今天伊斯兰教仍是这种状况），伍麦耶人建立了王朝，以大马士革为首都统治伊斯兰90年（公元661—750）。后来的一场革命带来朝代更替，伍麦耶人只保住了西班牙，在那里他们又统治了300年。

新王朝是从穆罕默德的伯父阿巴斯那里传承而来的，所以历史上它的统治者就是阿巴斯人。公元762年他们建立了新首都巴格达，为了获取建筑材料，他们洗劫了古巴比伦人和波斯人的遗迹。英语单词“代数”来自一本书的标题，这本书就是在公元820年阿巴斯王朝的巴格达所写的，写此书的人有一个滑稽的名字：Abu Ja’far Muhammad ibn Musa al-Khwarizmi[8]。我将像大家一样称其为花拉子米。



第5至第7阿巴斯哈里发王朝（也就是从公元786年到公元833年）统治下的巴格达是一个伟大的文化中心，现代西方人只模糊地知道这是一个《一千零一夜》中所描绘的有巫师、奴隶、大篷车和远途旅行商人的世界。阿拉伯人自己却认为那时的巴格达正处于黄金时代，尽管事实上阿巴斯哈里发已经不具有强大的军事力量来维持最初从伊斯兰人那里所赢得的各领地，而且正在因北非和高加索等地区造叛反而进一步失去领地。

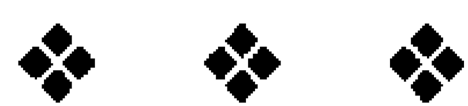
波斯是阿巴斯领土的一部分，在教会和现实权力两方面都受控于哈里发。然而，从1400年前的米底帝国开始，波斯就已经有了高度文明，而公元800年的阿拉伯人只是脱离荒芜沙漠家园的第六代。因此，在某种程度上阿巴斯王朝对波斯人有文化上的自卑情节，就如罗马人对希腊人一样。

波斯人之外是印度人，第一次伟大的穆斯林扩张没有触及他们。公元4世纪和5世纪，北印度已统一在笈多王朝的统治之下，而且自此之后北印度又渐渐分成各小州，这种情况一直持续到土耳其列强在10世纪末的入侵。中世纪的印度文明对一些数值特别着迷，尤其是一些大数值，他们还特意给这些数值起了名字（你也许曾经看到过梵文词汇塔拉克坎纳，它代表的数值是10^{53}）。这时终于有了数字零的发现，这一不朽的荣耀是属于印度人的，也许是属于公元598年到670年的数学家婆罗摩笈多的，此外我们所说的阿拉伯数字，实际是也是印度人发明的。

印度人之外的当然是中国人。至少从7世纪中期开始，印度人通过佛教僧侣唐玄奘开始与中国人接触，而波斯人经由丝绸之路与中国人进行频繁的贸易往来。中国人很早就有了自己的数学文化，我将在第9章对此加以介绍。

这样，悠闲且好奇心强烈的巴格达居民精通当时文明世界中所知道的任何事情。他们通过亚历山大城以及他们与拜占庭帝国之间的贸易往来了解希腊人和罗马人的文化，他们能够容易地接触到波斯、印度和中国的文化。

把阿巴斯王朝的巴格达变成一个保存丰富知识的理想中心所需要的一切就是一个学院，一个能够查询已成书的文献而且能够举行演讲和学术会议的地方。不久，这样的学院就出现了，人们称它为智慧之所（Dar al-Hikma）。这个学院最鼎盛的时期是第七代阿巴斯王朝阿里发马蒙统治的时期。用亨利·罗林森爵士的话说，马蒙统治时期的巴格达“在文学、艺术和科学等领域同西班牙的科尔多瓦一样达到了这个世界最高水平，而在商业和财富方面则远远超过了科尔多瓦。”花拉子米就生活和工作在这个时期。



我们对花拉子米的生活了解得很少。对他所生活的年代也只知道个大概。在伊斯兰历史学家和目录学家的著作中，关于花拉子米的生活有

一些零碎的干巴巴的记录条目，如果想了解得更详细，我建议读者参考科学传记辞典。我们只知道他编写的几本著作：一本是关于天文学的，一本是关于地理学的，一本是关于犹太历法的，一本是关于印度数字体系的著作，还有一本是编年史。

关于印度数字体系的这本著作只有拉丁语译本保存了下来，它的卷首语是“根据花拉子米……”（Dixit Algorismi…）。这本书展示了现代十进制算术数值体系的计算法则，这些法则是印度人发明的，它的影响非常大。因为这些卷首语，掌握了这种“新算术”（相对于旧罗马数字体系，罗马体系对运算是毫无价值的）的中世纪欧洲学者把自己称为“算术家”（algorismist）。后来人们用“算法”（algorithm）一词来表示用有限次确定步骤可以完成的任何计算过程。这是现代数学家和计算机科学家使用的含义。

这本书真正吸引我们的是它的标题《利用还原与化简运算的简明手册》。这是一本代数和算法教科书，是自600年前丢番图的《算术》之后这个领域第一本意义重大的著作。此书分三部分，其话题分别是二次方程求解、面积和体积测量、研究复杂的伊斯兰继承法所需要的数学。

严格说来，这三部分中仅第一部分是代数，这有些令人失望。而且花拉子米没有文字符号体系，因此也就没有用字母和数值表示方程的方法，也没有表示未知量和未知量的幂的符号。对于我们所写的方程

$$x^2+10x=39$$

丢番图把它写成

$$\Delta^{Y}\overline{\alpha}\varsigma\overline{\iota}\,'\iota\sigma M\overline{\lambda\theta}$$

而它在花拉子米的著作中的出现形式则是：

> 一个平方与10个该平方的根之和等于39卡塔尔。也就是说，当一个平方加上它自己的根的10倍后总和等于39时，这个平方是什么？

[卡塔尔（dirhem）是一种货币单位。花拉子米利用它表示我们今天所说的常数项，即x^0的项。]

另外，丢番图从几何方法向符号处理的历史性转变在花拉子米的著作中还是没有出现。这并不太令人惊讶，因为他没有要处理的符号，但这与600年前丢番图的伟大突破相比有一些倒退。范德瓦尔登说："我们应该可以排除花拉子米的工作深受古典希腊数学影响这种可能性。"

事实上，花拉子米主要的代数成就是提出了把方程作为研究对象的想法，他对所有一个未知量的一次和二次方程进行了分类，并给出了处理它们的法则。他把这些方程分成6个基本类型，我们把它们用现代符号体系写出来，就是：

(1) $ax^2 = bx$　　(3) $ax = b$　　(5) $ax^2 + c = bx$

(2) $ax^2 = b$　　(4) $ax^2 + bx = c$　　(6) $ax^2 = bx + c$

对我们来说，上面某些方程显然是同一类型的，这是因为我们有了负数的概念，而花拉子米却没有这样的知识。当然他可能会提到减法，提到一个量超过另一个量，或者一个量比另一个量少，但是，他的自然算术意识是用正量看待每一件事情。

至于说到处理技巧，就是"还原"（al-jabr）和"化简"（al-muqabala）的引入。如果有下面的方程：

$$x^2 = 40x - 4x^2$$

（或者如花拉子米所说："一个平方等于四十个该平方的根少四个平方。"），你如何把这个方程处理成6个标准类型中的一个呢？"还原"这个方程：把这个方程的两边加上$4x^2$，于是就得到类型(1)的方程：

$$5x^2 = 40x$$

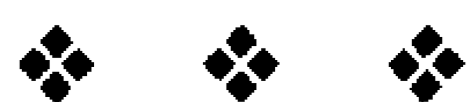
这一做法是在方程两边加上相等的项。另一种相反的做法是从方程两边减去相等的项，即"化简"。例如，把下面的方程：

$$50 + x^2 = 29 + 10x$$

变成类型(5)的形式：

$$21+x^2=10x$$

就是从方程两边减去29。



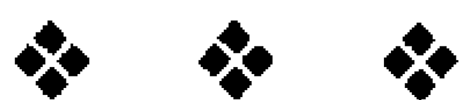
这两种处理方法都不是新的。事实上，还原和化简在丢番图时代就已经发现，当然二者都有很丰富的文字符号体系来帮助处理。图默在科学传记辞典中说："花拉子米的科学成就其实很平庸，但是它们的影响力是巨大的。"

事实上，我担心此刻读者容易产生这样的想法：这些古代和中世纪的代数学家"不是非常聪明"。我们说在公元前1800年巴比伦人开始对写成文字问题的二次方程求解，而2600年后的花拉子米仍在对写成文字问题的二次方程求解。

我承认这的确有点令人郁闷。然而某种程度上这还是令人振奋的。形成符号代数的进步极其缓慢，证明了这一课题的层次非常高。借用约翰逊博士的比喻，奇妙之处不在于我们花如此长的时间去了解这件事怎么做，而在于我们一直在这么做。

而事实上，代数的发展到了中世纪[9]中期才开始有了一点起色。花拉子米之后的著名数学家掘起于东西方的穆斯林大地。撒彼特·伊恩·克拉就是花拉子米之后的一代人，他也生活在巴格达，在天文数学和纯数论方面做出了令人瞩目的工作。一个半世纪之后，西班牙的穆斯林聚居地科尔多瓦城的默哈麦德·贾亚尼写了第一篇关于球面三角学的论文。然而，这些成就都没有使代数取得大的进步。而且没有人尝试去复制丢番图在文字符号体系领域的伟大突破。所有人都使用文字表述他们的问题。

我将仅对另一名中世纪伊斯兰的数学家加以详细的介绍，不仅因为他值得介绍，而且他还是与文艺复兴初期的欧洲接轨的桥梁，代数只有到了文艺复兴时期才真正开始有了起色。



奥马·海亚姆作为《鲁拜集》的作者闻名于西方。这是一本四行诗集，展示了极端个人的人生观。它以一个酒鬼为线索展示死亡威胁享乐主义的人生观点，某种程度上以霍斯曼为原型。爱德华·菲茨杰拉德把其中的75首诗翻译成英语四行诗，每一首的押韵方式是a-a-b-a。菲茨杰拉德的《奥马·海亚姆的鲁拜集》于1859年出版，在第一次世界大战前的英语国家里非常受欢迎。（一本带有华美珠宝封皮的原本同泰坦尼克号一起沉入了大海。）

科学传记辞典认为海亚姆最有可能的生存年代是1048年到1131年，我也不得不使用这一年代推测。这样一来，海亚姆至少晚于花拉子米250年。在审视中世纪的文化活动时，一定要牢记这一巨大的时间跨度。

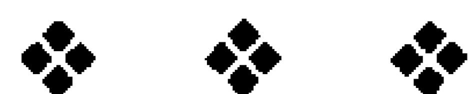
海亚姆生活和工作的地区位于伊斯兰第一次大征服区的东边。最东边的区域包含美索不达米亚、今天的伊朗北部地区和中亚的南部地区（今天的土库曼斯坦、乌兹别克斯坦、塔吉克斯坦和阿富汗）。在海亚姆时代，这里是既有种族冲突也有宗教冲突的地区。其中主要的种族有波斯人、阿拉伯人和土耳其人。宗教冲突都发生在伊斯兰教内部：首先是逊尼派和什叶派间的冲突，其次是什叶派中两个派系——主体教派与分裂出来的伊斯玛仪教派[10]之间的冲突。

土耳其人最初是从遥远的中亚人分裂出来的游牧民，他们被日渐衰败的阿巴斯王朝雇用，充当雇佣兵。当然，土耳其人不久就认识到了在这种关系中怎样取得真正的权力平衡，除了9世纪末短暂的复苏之外，后来的阿巴斯王朝的哈里发都是这些土耳其军队的傀儡。他们仅有的安慰是这些土耳其人都皈依到正统的伊斯兰教（也就是逊尼派）了。美索不达米亚再往东的领地由原来的阿巴斯王朝失陷给10世纪兴起的波斯王朝（和什叶派）。这些短命的波斯王朝与阿巴斯王朝一样雇用土耳其军队。当时一位土耳其将军推翻了波斯统治，建立了伽色尼王朝，这是第一个土耳其帝国。由于土耳其人很聪明地沿用了已被征服的波斯人的国家管理体制和高度发达的文明，他们的统治相当了不起，吸引了著名的波斯

学者和诗人。他们还发动了几次向南亚的进攻，把伊斯兰教带入了如今的巴基斯坦和印度地区。

1037年，海亚姆出生的几年前，一位名叫塞尔柱的伽色尼土耳其雇佣兵造反并打败了伽色尼军队。这个新建立的土耳其政权迅速扩张。1055年，海亚姆只有7岁，塞尔柱的孙子占领了巴格达，并自封为苏丹，意思就是“统治者”。这一切表明，此时的哈里发只是宗教领袖，就如罗马教皇一样。

塞尔柱王朝从11世纪后半期开始到12世纪大部分时期统治着伊斯兰的东部领土。他们的领土一直向西扩张到了圣地①和埃及的边界（当时埃及等地由什叶派统治，他们属于伊斯玛仪派，而塞尔柱人是正统的逊尼派）。正是这个塞尔柱在1071年的曼齐克特战役中打败了拜占庭帝国，占领了安纳托利亚，奠定了现代土耳其的第一个根据地。正是由于失去了安纳托利亚，拜占庭不得不求助于西欧人的帮助，因而组建了十字军。十字军艰苦跋涉穿越安纳托利亚到达圣地时，在安提俄克和耶路撒冷要面对的人同样是塞尔柱土耳其人。



因此，海亚姆在塞尔柱土耳其的统治之下度过了他的一生。他的重要赞助人是第三代塞尔柱苏丹马里克·沙[11]，后者统治时期是1073年到1092年，首都是今天伊朗境内的伊斯法罕市，位于伊拉克巴格达以东700公里。马里克·沙没有他的大臣尼珊·阿穆鲁克有名，这位名臣是历史上有名的国家管理者，也是一位外交天才。阿穆鲁克同海亚姆一样是波斯人。据说他们两人时常与哈桑·萨巴赫在一起，这个人是暗杀派[12]的创始人，他们是当时的“波斯三巨头”，是塞尔柱帝国的三个重要人物。

在宗教方面，马里克·沙宫廷似乎很有包容心，中世纪伊斯兰教就是在这样的环境下发展起来的。这也许非常适合海亚姆。他的诗展示出

① Holy Land，指以耶路撒冷为中心的周边地区。——编者注

怀疑论和对人生的不可知的态度，因此他的同时代人通常认为他是自由思想家。作为伊斯法罕的大天文台台长，海亚姆主要担心无法写作和学习，所以他尽量不参与麻烦的事情，只是按要求编写传统宗教手册或者是主持去麦加的朝圣（这是每个穆斯林的职责）。从现存的这些诗和传记事实可以看到，海亚姆因其相当的亲和力打动了现代的读者。

作为代数学家，他的主要影响是在他二十几岁去伊斯法罕之前写的一本书，书名是《还原和化简问题示范》。

同花拉子米和其他所有中世纪穆斯林数学家一样，海亚姆忽视了或者说不知道丢番图在文字符号体系进程中的伟大突破。他是用文字阐述每一件事物的。另外也同较早的希腊人一样，他用了很强大的几何方法，能够很自然地把数值问题的求解转化成几何方法。

海亚姆对代数发展的主要贡献是他首先开始认真研究三次方程。由于缺少适当的符号体系，并且很明显不愿意采用负数，海亚姆的研究显得很吃力。例如，我们这样书写的方程$x^3+ax=b$，海亚姆将其表述为“一个立方加上几条边等于一个数”。但是他还是提出并解决了涉及三次方程的若干问题，只是他的解决方法都是几何方法。

这不是三次方程的第一次露面。我们已经说过，丢番图对此也做过一些尝试。甚至在这之前，阿基米德在考虑如何把一个球分成两部分，使得它们的体积比是给定的比例等这一类问题[13]时，也曾经研究过三次方程。（如果进一步思考一下，那么你就会想到这与阿基米德对浮体的兴趣有关。）海亚姆似乎是第一个把三次方程看成不同类型问题的人，而且还把它们分成14个类型，他十分清楚如何使用几何手段解决这些类型。

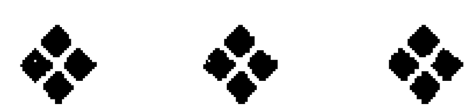
下面是海亚姆把这一类型问题简化成一个三次方程的例子。

> 画一个直角三角形。从直角做这个三角形斜边上的垂线。如果这条垂线的长度加上这个直角三角形最短边的长度等于斜边的长度，你能说出这个直角三角形是什么样的直角三角形吗？

答案是这个三角形最短边与第二短边的比必须满足下面的三次方程，这个比完全决定了这个三角形是什么样的直角三角形。

$$2x^3-2x^2+2x-1=0$$

这个方程唯一的实数解是0.647 798 871…，这是一个非常接近于有理数103/159的无理数。所以短边分别是103和159的直角三角形非常接近这个问题的答案，读者自己可以很容易验证。[14]海亚姆采用了一个间接的方法，他利用一个略微不同的三次方程，而且可以通过两个经典的几何曲线的交点来求解这个三次方程。



下面列出到目前为止各事件的简略概要。

- 古巴比伦人发明了一些技术，用以在有限范围内对含一个未知量的一次方程和二次方程求解。
- 后来，古希腊人用几何方法解决了类似的方程。
- 在公元3世纪，丢番图拓宽了研究范围，研究了很多其他类型的方程，其中包括高阶方程、多变量方程以及同类方程的联立方程组。他还为代数问题开发了史上第一套文字符号体系。
- 中世纪伊斯兰学者为我们发明了"代数"一词。他们开始把方程作为有价值的研究对象专门进行研究，同时根据利用已有技术解决它们的难易程度对线性方程、二次方程和三次方程进行了分类。

在介绍海亚姆时，我提到了可怕的曼齐克特战役，随后就是东基督教徒的大撤退，以及之后奇怪、混乱且一直存在争议的反攻：十字军东征。在这段时间里（即海亚姆生活的年代），西欧文化在黑暗时代之后的困境中崛起。希望之灯首先在意大利燃起，这里的灯也最明亮。也就是在这里，我们将遇到后面几名代数学家。

注　解

[1] 新柏拉图派哲学的创始人柏罗丁是另一位亚历山大人，他很有可能与丢

番图是同一时代的人。伯特兰·罗素曾说："在世俗意义下的不幸之人却毅然决定去寻找理论世界的最高幸福，柏罗丁就是这类人的先驱。"新柏拉图派哲学认为数学至高无尚，旧柏拉图哲学和原始柏拉图哲学也都这样认为。后来的一位新柏拉图派哲学家马林努斯评论说："我希望所有事情都是数学。"

[2] 犹太人也许已经回到亚历山大城，因为公元640年这个城市的穆斯林统治者报导说，这个城市有"四千名附属的犹太人"。

[3] 吉本在《罗马帝国兴衰史》一书的第47章中写到："恕我不了解实情，但这些暗杀者也许没有留意他们的牺牲品是否还活着。"因《水孩》一书而闻名的查尔斯·金斯莱因专门写了一本关于海帕希亚的小说，其中记载，在被用牡蛎贝壳刮骨后，她仍然活着。这本小说和受到其启发而做的查尔斯·威廉·米契尔的画一样，完全是夸张的维多利亚式风格。

[4] 本书中"波斯人"的涵义较广，泛指现今的讲印欧语系的伊朗和南中亚的人们（除亚美尼亚人）。这里实在没有太合适的词了，Aryan（雅利安人）这个词贬义的会令人不悦，"伊朗人"指现代的伊朗公民，而不是指说印欧语系的一些人。有很多人不高兴自己被说成是"波斯人"，在不同的历史背景下，这个词容易引起混淆。我真可怜呐，必须尽全力才能解释清楚。

[5] 吉本说赫拉克利乌斯于亚历山大城沦陷50天后死于"水肿"。

[6] 在吉本的书中这个人的英文名字为Amrou，而不是Amribn al'As。

[7] 基督一性说认为，救世主的人性和神性实际上是一种东西。涅斯特利教派则持反对意见。451年由迦克墩会议采纳并由基督主要教堂保持下来的传统提法是耶稣基督人性和神性是一件事情，同时也是两件事情，"没有混淆，没有变化，不会分割，不会分离的两种天性"。

[8] 意思是"Ja'far的父亲Mohammed，Musa的儿子Khwarizmian"。Khwarizm是古代的一个洲，也就是现在的乌兹别克斯坦。顺便提一下，以"al-"开头的阿拉伯名字总以它们的第二部分编入索引和目录。例如Al-Khwarizmi在DSB中属于K字头索引，而不属于A字头索引。

[9] 中世纪开始于公元476年9月4日星期六，当时最后一位西罗马皇帝被奥多亚克蛮族人废除。结束于公元1453年3月29日星期二，君士坦丁堡失陷。如果我的数字正确，中世纪精确的中间点大约是公元965年6月15日星期天的午夜。

[10] 在英语中有时候把什叶派的两部分分别称为"十二"什叶派和"七"什叶派。什叶派相信穆罕默德的堂兄和女婿的第四任哈里发阿里是第一位教长。这是有无限权力的宗教头衔，他下面还有几位教长。教长头衔是由父亲传给儿子的。然而，这一继承线路被打破了，什叶派最终分裂为以第七任教长伊斯梅尔为首的

一个分支和以第十二任教长玛迪为首的一个分支。伊斯玛仪派是“七”什叶派，实际上现在大部分什叶派都属于“十二”什叶派。

[11] 马里克·沙（Malik Shah）的名字传递了一些塞尔柱帝国中种族平衡的信息。Malik和Shah分别是“国王”一词的阿拉伯语和波斯语。当然，马里克·沙实际上是土耳其人，是三个种族的混血儿。

[12] （Assassins）这一教派属于伊斯玛仪派，因此，与塞尔柱帝国的逊尼派教徒统治者和其他什叶教徒发生冲突。因为暗杀派把很多古波斯人的信条中的神秘方法用于伊斯兰教。这一做法遭到所有人的反对，暗杀派最终撤退到伊朗北部的偏远山区，在那里他们开始实施恐怖的政治暗杀方案。（非常了解他们的十字军战士称哈桑· 萨巴赫为“山中老人”。）撇开暗杀不谈，历史学家对通常被称为Assassins的行动含义也有争议。例如，没有证据显示他们对对手使用大麻，但出于宗教的目的他们自己经常使用这种麻药。

[13] 利用现代术语描述这一问题：假设处于平坦水平面上的一个直径为D的球，被一个高度为x的平行水平平面在它的顶部切下一个薄片，这样，这个球的剩余部分是原来球体体积的百分之R。x应该是多少呢？解下面的三次方程就可以求得答案：

$$2\left(\frac{x}{D}\right)^3-3\left(\frac{x}{D}\right)^2+\frac{R}{100}=0$$

例如，如果R是50%，那么显然x等于D的一半（即$x/D=1/2$），因为有

$$2\left(\frac{1}{8}\right)-3\left(\frac{1}{4}\right)+\frac{50}{100}=0$$

[14] 根据毕达哥拉斯定理，短边分别为103和159的直角三角形的斜边长$\sqrt{103^2+159^2}$，这个值是189.446 562 386 55…，其垂线长度是86.446 540 880 49…，所以，如果把这个数加上103，就会得到斜边长的近似值。

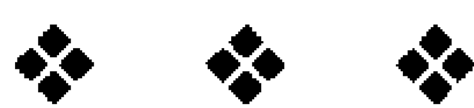
数学知识

三次方程和四次方程

中世纪之后，代数的第一次伟大进步是得出了三次方程的一般解法，紧随其后的就是四次方程的一般解法。我将在第4章用16世纪的观点来讲述这段历史，这里只是对基础代数进行简单的说明，以便阐明这个话题及其困难之处。

首先需要弄清楚我们寻找的对象。对于三次方程或者任意阶方程，寻找达到要求精度的近似数值解不是一件很难的事情。有时都能猜测出一个正确的解。画出图来就更容易找到这个解了。希腊人、阿拉伯人和中国人都掌握了很多成熟的算术和几何方法。中世纪欧洲数学家也熟悉这些方法，而且还能够很精确地计算出一个三次方程真解的数值解。

他们没有得到的是真正的代数解，即求解三次方程的一般公式，就像第1章的注解[6]给出的二次方程的求解公式一样。那种一般形式的公式是近代数学家找到的。只有找到这样的公式才能认为三次方程是可解的。



下面是含一个未知数的三次方程的一般形式[1]：$x^3+Px^2+Qx+R=0$。首先能做的就是消去包含x^2的项，做下面这个简单的代数变换即可：

$$x^3+Px^2+Qx+R=\left(x+\frac{P}{3}\right)^3+\left(Q-\frac{P^2}{3}\right)\left(x+\frac{P}{3}\right)+\left(R-\frac{QP}{3}+\frac{2P^3}{27}\right)$$

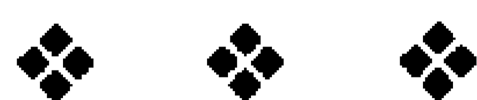
用另外一种方式表示它：

$$x^3+Px^2+Qx+R=X^3+\left(Q-\frac{P^2}{3}\right)X+\left(R-\frac{QP}{3}+\frac{2P^3}{27}\right)$$

其中$X=x+P/3$。所以，只需要通过代数替换就可以将任意的三次方程转化成没有x^2项的方程。如果能够通过求解X解决这个较简单的方程，那么就能够通过减去$P/3$而得到x。没有x^2项的这种三次方程有一个更独特的专业名字：不完全三次方程[2]。

概括地说，我们只需考虑下面这样的不完全三次方程：

$$x^3+px+q=0$$



到此一切顺利，但是不完全三次方程的一般求解公式是什么呢？回想一下一般的二次方程

$$x^2+px+q=0$$

它有两个解

$$x=\frac{-p+\sqrt{p^2-4q}}{2} \text{和} \ x=\frac{-p-\sqrt{p^2-4q}}{2}$$

因为负数没有实平方根，所以如果p^2-4q是负数，那么上面的根就是复数。如果p^2-4q正好等于零，那么上面两个根就相等，因此从实数观点看只有一个根。当然，如果p^2-4q是正数，那么就有两个实数根。

上述这些情况都可以用二次抛物线说明。如果画出$y=x^2+px+q$的二次抛物线，那么这个二次方程的根就是$y=0$的地方，即这条曲线与水平轴相交的地方。上面我总结的三种情况——没有实根、只有一个实根及有两个实根，分别如图CQ-1、图CQ-2及图CQ-3所示。

为了搞清楚从三次方程出发会得到什么样的一般求解公式，我们可以从绘图开始。存在三种基本情况，如图CQ-4、图CQ-5和图CQ-6所示。注意，所有的三次曲线都是从第三象限开始，到第一象限结束的。这是因为，当x非常大的时候（正的或者负的），x^3+px+q中的x^3项“淹没了”

其他项。对于“足够大”的x值，x^3+px+q就会像x^3一样。（“足够大”的精确尺度取决于p和q的大小。）现在，根据符号法则，负数的立方是负的，正数的立方是正的，所以三次多项式曲线走向总是从第三象限走向第一象限。那么，因为这条曲线必然与水平轴相交于某个位置，所以x^3+px+q一定至少有一个实根。

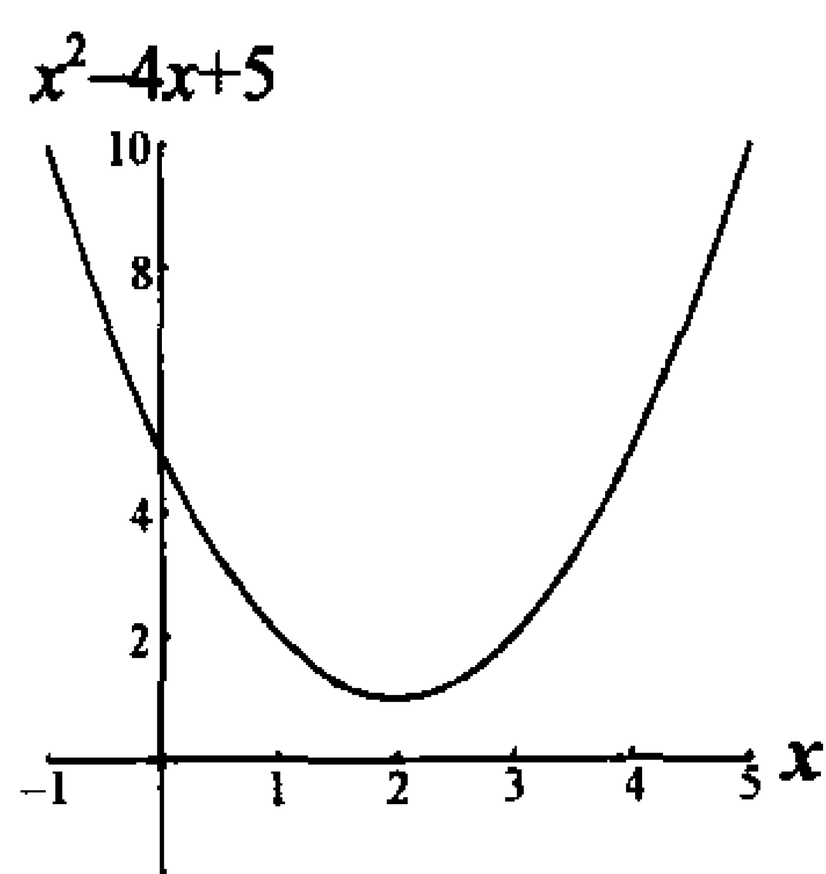


图　CQ-1

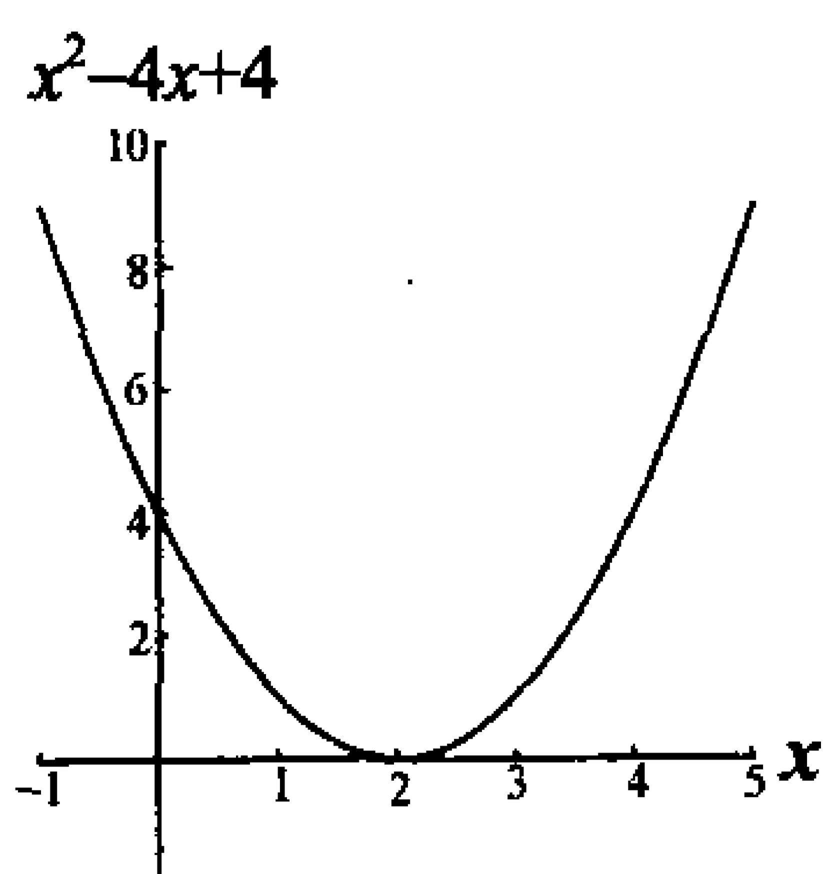


图　CQ-2

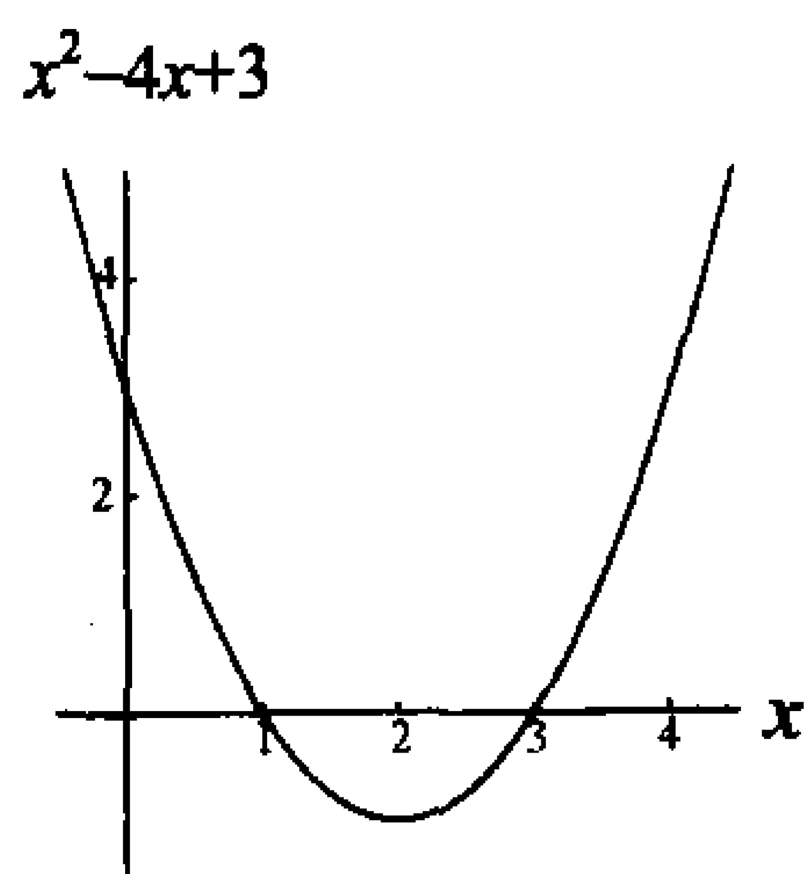


图　CQ-3

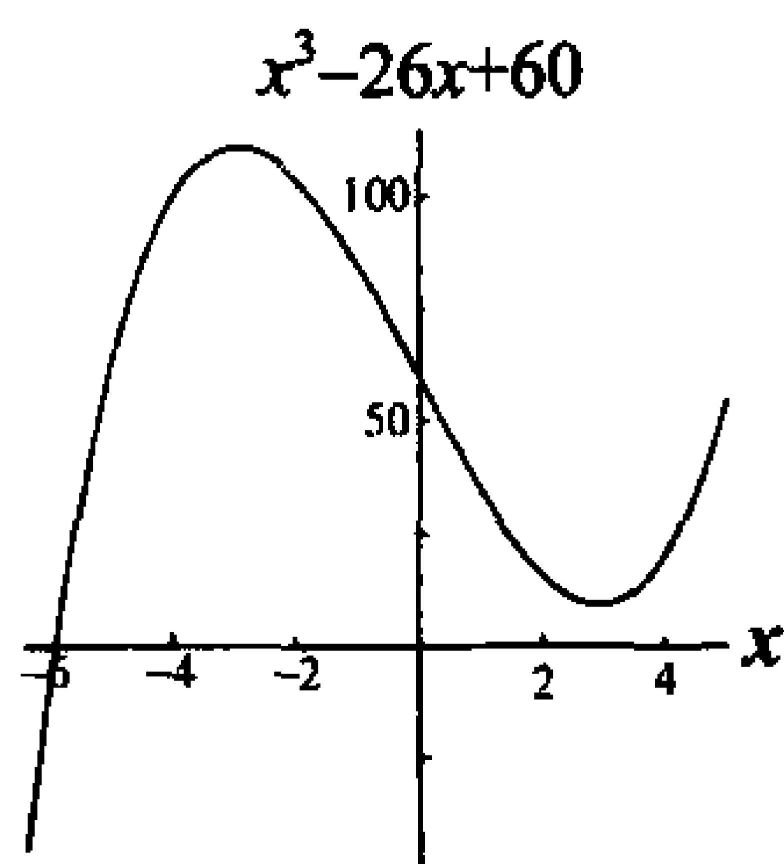


图　CQ-4

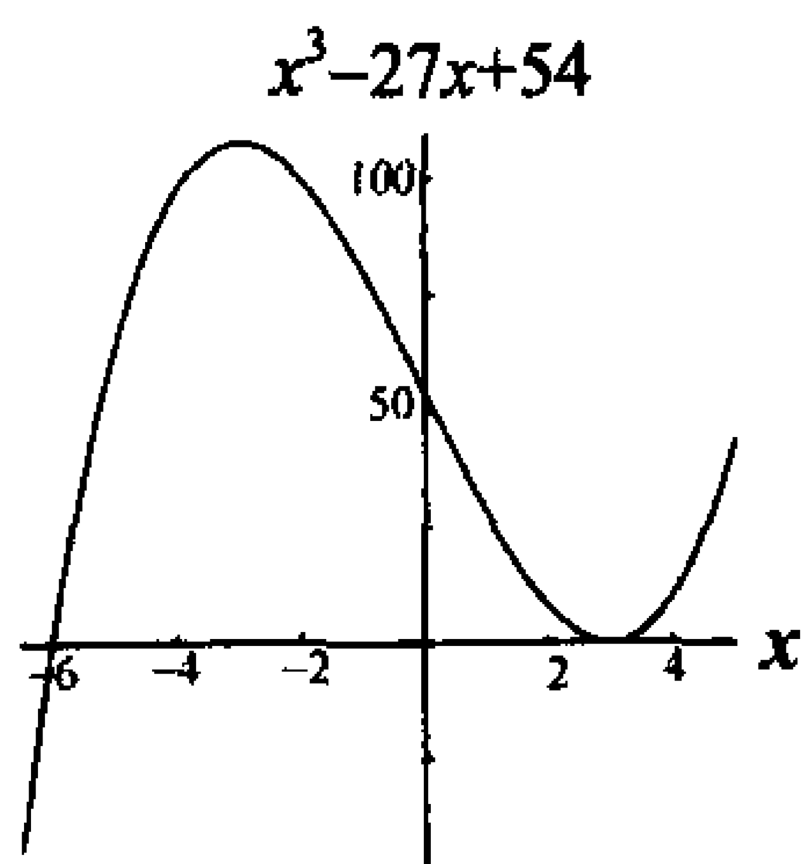


图　CQ-5

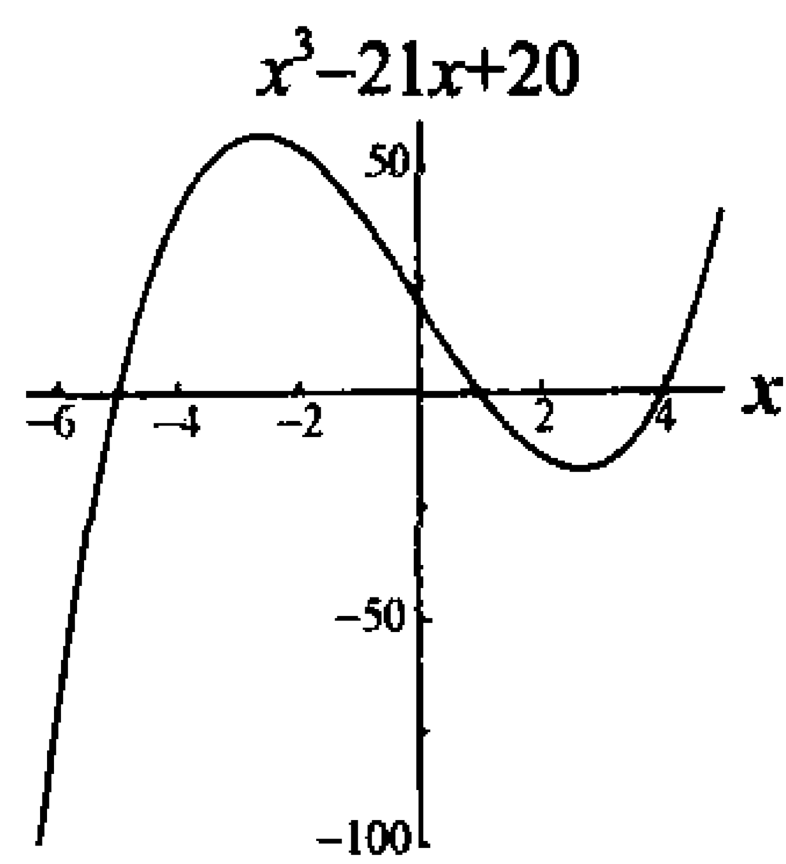
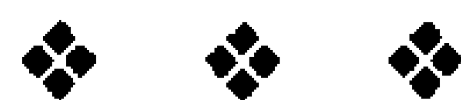


图　CQ-6

正如这些图所展示的那样，事实上，一个三次方程可能有一个、两个或三个实根，但不能一个也没有，而且也不能有四个或更多的根。根据对二次方程的了解可以猜测，对于大多数只有一个实根的情况，存在两个复根，只是图上看不出来。这个猜测是正确的。



事实上，下面是三次方程$x^3+px+q=0$的三个一般解：

$$x=\sqrt[3]{\frac{-q+\sqrt{q^2+\left(4p^3/27\right)}}{2}}+\sqrt[3]{\frac{-q-\sqrt{q^2+\left(4p^3/27\right)}}{2}}$$

$$x=\left(\frac{-1+\mathrm{i}\sqrt{3}}{2}\right)\sqrt[3]{\frac{-q+\sqrt{q^2+\left(4p^3/27\right)}}{2}}+\left(\frac{-1-\mathrm{i}\sqrt{3}}{2}\right)\sqrt[3]{\frac{-q-\sqrt{q^2+\left(4p^3/27\right)}}{2}}$$

$$x=\left(\frac{-1-\mathrm{i}\sqrt{3}}{2}\right)\sqrt[3]{\frac{-q+\sqrt{q^2+\left(4p^3/27\right)}}{2}}+\left(\frac{-1+\mathrm{i}\sqrt{3}}{2}\right)\sqrt[3]{\frac{-q-\sqrt{q^2+\left(4p^3/27\right)}}{2}}$$

需要稍微解释一下这几个公式。哦，不，得详细解释才行。

首先，注意到每个公式中立方根下的两个项的差异只是符号不同，一个是加，一个是减。如果再仔细观察一下，你会发现它们是下面这个二次方程的两个根：

$$t^2+qt-\frac{p^3}{27}=0$$

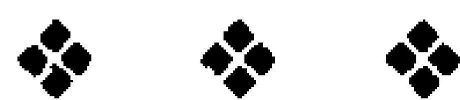
（我把x换成t作为“虚拟”变量，以避免两个变量混淆。）为方便起见，我们把这两个根称为α和β。

其次，第二行和第三行中的那两个复数是什么呢？它们是1的立方根。当然，在实数范围内，数1只有一个立方根，就是它本身，因为1×1×1=1。然而，在复数范围内，它有三个根。如果你提前预习了第6章和第7章之间穿插的关于单位根的知识就会知道，上面出现的第一个复数通常用ω来表示，而第二个复数用ω^2来表示，它们互为对方的平方。

现在，我可以非常简洁地写出一般三次方程的解。它们是：

$$x=\sqrt[3]{\alpha}+\sqrt[3]{\beta}$$
$$x=\omega\sqrt[3]{\alpha}+\omega^2\sqrt[3]{\beta}$$
$$x=\omega^2\sqrt[3]{\alpha}+\omega\sqrt[3]{\beta}$$

其中，ω和ω^2是1的复数立方根，而α和β是二次方程$t^2+qt-(p^3/27)=0$的根。



再仔细看一下上面的三个解，似乎第一个是实数，第二个和第三个

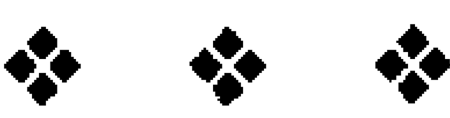
是复数。既然三次方程能够有三个实根（如图CQ-6所示），这个方程的根怎么会这样呢？

问题在于α和β本身都可能是复数。它们两个都包含$[q^2+(4p^3/27)]$的平方根，而这个数可能是负的。如果它是负的，那么α和β就都是复数。如果它是正的，那么α和β就都是实数。它们都是复数的情况就是所谓的不可约情况[3]。那么你就不得不去寻找一个复数的立方根，这可不是一件容易事。因此，三次方程的一般解尽管在理论上非常令人满意，但不是非常实用。

我将这三个根的实际情况概括成下面的表格，它们与$[q^2+(4p^3/27)]$的符号有关。我把$[q^2+(4p^3/27)]$称为这个方程的判别式。表中最下面一行表示不可约情况。

	第一个根	第二个根	第三个根
判别式为正	实数	复数	复数
判别式为0	实数	相等实数	
判别式为负	实数	实数	实数

如果这个判别式为零，那么α和β相等，所以三个根归结为$2\sqrt[3]{\alpha}$、$(\omega+\omega^2)\sqrt[3]{\alpha}$和$(\omega^2+\omega)\sqrt[3]{\alpha}$。显然后两个根相等，很容易就可以验证$(\omega+\omega^2)$等于−1。



我已经不加证明就给出了三次方程的一般求解公式。在现代符号体系下证明并不难。为了求解$x^3+px+q=0$，首先用两个数的和表示x：$x=u+v$。当然，表示x的方法有无穷种。这时，这个三次方程就变成了$(u+v)^3+p(u+v)+q=0$。然后，再重新整理成如下形式：

$$\left(u^3+v^3\right)+3uv(u+v)=-q-p(u+v)$$

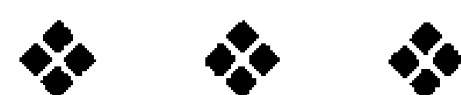
因为u和v可以是任意数，所以，如果我选出特殊值，满足

$$u^3+v^3=-q \text{ 和 } 3uv=-p$$

那么我就得到一个解。对于上面的两个方程，根据第二个方程用u和p表示v，然后再把它带回到第一个方程中，于是得到下面的方程：

$$u^6 + qu^3 - \frac{p^3}{27} = 0$$

这是关于未知数u^3的二次方程。因为我同样可以很容易地表示u而不是v，这样就可以得到一个以v为变量的相同的方程，u^3和v^3是上面这个二次方程的解。那么u和v就是这些解的立方根，或者是这些立方根再乘以1的立方根。所以可能的解是$u+v$、$\omega u+\omega^2 v$和$\omega^2 u+\omega v$。（注意，类似$u+\omega v$这样的组合不可能是解，因为有条件$3uv=-p$，表明加号两边的数乘在一起必须得到实数。当u和v是实数时，ωu乘以$\omega^2 v$是一个实数，因为$\omega^3=1$。）



一般四次方程

$$ax^4 + bx^3 + cx^2 + fx + g = 0$$

的求解方法总体说来就是降阶法。同前文的做法一样，我们通常将上面的方程简化或者“降低阶次”，变为下面的形式：

$$x^4 + px^2 + qx + r = 0$$

直接进行替换，就可以把这个方程写成两个平方表达式的差。在这些方程的求解过程中，你最终会得到一个三次方程：与把三次方程化简成二次方程的方法类似，可以把一个四次方程化简成一个三次方程。

这四个解是用包括p、q和r的平方根和立方根表示的表达式。把这四个解完整地写出来需要占用大量篇幅，所以这里只写出其中的一个：

$$x = \frac{1}{2}\sqrt{-\frac{2p}{3} + \frac{\sqrt[3]{2}t}{3\sqrt[3]{u}} + \frac{\sqrt[3]{u}}{3\sqrt[3]{2}}} - \frac{1}{2}\sqrt{-\frac{4p}{3} - \frac{\sqrt[3]{2}t}{3\sqrt[3]{u}} - \frac{\sqrt[3]{u}}{3\sqrt[3]{2}} - \frac{2q}{\sqrt{-\frac{2p}{3} + \frac{\sqrt[3]{2}t}{3\sqrt[3]{u}} + \frac{\sqrt[3]{u}}{3\sqrt[3]{2}}}}}$$

其中，$t=p^2+12r$且

$$u = 2p^3 + 27q^2 - 72pr + \sqrt{\left(2p^3 + 27q^2 - 72pr\right)^2 - 4t^3}$$

这看起来好复杂，但是它只包含算术运算，涉及平方根、立方根以及p、q和r的组合。我们已经看到二次方程的一般解和三次方程的一般解在复杂度上有所增加，这又在类似量级上增加了复杂度。

你可能由此推测五次方程的一般解应该比这更复杂，可能需要整整一页甚至更多页，但它的解也只是由平方根、立方根，可能还有五次方根以不同的方式互相嵌套而成的。有了求解二次方程、三次方程和四次方程的经验，这一推测完全是合理的。但是，这是不正确的。

注 解

[1] 注意，我假设x^3的系数是1。这样假设没有失去一般性。更一般的形式是$ax^3+bx^2+cx+d=0$。a可以是零或其他数。如果它是零，这个方程就不是三次方程。如果它不是零，我可以对它做除法，把x^3的系数化简成1。

[2] 现在常用的术语是“简化的三次方程”，但我更喜欢这个旧的术语。

[3] 这是一个非常容易引起混淆的术语，最好是限定在这一历史背景之下使用。在更一般的方程理论中，一个不可约方程是这样的方程：如果不扩大数域，引入一个新的“俄罗斯娃娃”，那么它不能被因式分解。（详见“数学知识：域论”。）对于三次方程$x^3-7x+6=0$，其$q^3+4p^3/27$等于$-400/27$，所以这是“不可约情况”。然而多项式x^3-7x+6可以完美地因式分解成$(x-1)(x-2)(x+3)$，所以精确意义下，它不是不可约的。只是中间过程中的二次方程是不可约的。

第4章

商业与竞争

詹姆斯·沃尔什所著的《十三世纪，最伟大的世纪》于1907年出版，这是一部能激发人的求知欲的历史著作。书中大部分内容是为罗马天主教的辩护（沃尔什是福特汉姆大学教授，而福特汉姆大学是纽约城的一个耶稣会士机构），但是作者却选择了一个很好的角度，他从13世纪的社会进步、文化成就和恢复学习古典知识展开论述，仅仅是这些成就就让人无法轻视那个年代。伟大的哥特式大教堂，早期的大学，契马布埃和乔托，圣弗朗西斯和阿奎奈，但丁和《玫瑰传奇》，路易九世，爱德华一世，弗雷德里克二世，大宪章和工会，马可波罗和鄂多立克（他好像到过中国的拉萨）……在13世纪，很多方面都取得了巨大进步。比萨的列奥纳多（即我们常称的斐波那契）就活跃在这个世纪之初。

斐波那契是广大非数学领域人士熟知的一位数学名人，其原因是广为人知的斐波那契数列：

$$1,1,2,3,5,8,13,21,34,55,89,144,233,377,\ldots$$

这个数列中的每一项都是其前两项之和，如89=34+55。这个数列在当前非常热门的整数数列在线百科全书的编号是A000045。其数学和科学内涵非常丰富，因而有一本专门研究它的杂志《斐波那契季刊》。2005年8月的季刊包括这样的论文“通过超几何函数得出的斐波那契数列p进制内插值”。

对于非数学人士来说，这听起来有点吃惊，但实际上很容易证明[1]

这个数列的第n项正好等于

$$\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

例如，如果n等于4，对上式使用二项式定理[2]得：

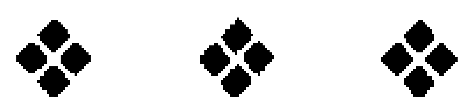
$$\frac{1}{\sqrt{5}}\left[\left(\frac{1+4\times\sqrt{5}+6\times5+4\times5\sqrt{5}+25}{16}\right)-\left(\frac{1-4\times\sqrt{5}+6\times5-4\times5\sqrt{5}+25}{16}\right)\right]$$

很容易就可算出上面的式子等于3。

斐波那契数列第一次出现在一本名为《算盘书》[3]的书中，这本书是由比萨的列奥纳多编写的，其内容是关于兔子数量的问题。

> 开始时只有一对兔子，如果每月每一对兔子繁殖一对新兔子，而这对新兔子第二个月开始就可以繁殖，而且没有死亡，那么一年内将有多少对兔子？

对于这个问题，我曾经能够想到的唯一方法是把这些月份标为A、B、C、D等。开始时的这对兔子记为A。第二个月我们仍有这对兔子A，而且还有它们生下的兔子对AB，总共两对兔子。在第三个月，兔子对A繁殖出另一对兔子AC。兔子对AB已经存在，但它们还没有开始繁殖，因而总共的兔子数目是三对。在第四个月，即月份D，兔子对A又生下一对兔子AD。此时，我们兔子对AB也开始繁殖出自己的一对兔子ABD。兔子对AC现在还不能繁殖。于是，A、AD、AB、ABD和AC是第四个月时所有的兔子对，依此类推。

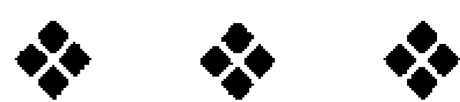


在《算盘书》的前言中，作者简单介绍了自己当时的生活。现代感十足的“名字加姓”在西欧还没有流行，所以人们只知道作者是比萨的列奥纳多，有时也被意大利化为列奥纳多·比萨诺。他属于一个有自己名字的家族，它的名字是Bonacci（波那契），所以他就用fi’Bonacci（波那契的儿子）来称呼自己，这个名字一直沿用到今天。

列奥纳多1170年左右[4]出生于比萨，这是波那契家族所在地。尽管比萨四周都是日耳曼帝国（即神圣罗马帝国）的领土，但是当时它是一个独立的共和国。列奥纳多的父亲是这个共和国的一名官员。大约1192年，他奉命代表比萨商人与当时位于北非海岸的布吉亚[5]殖民地进行贸易往来。不久后他就带着列奥纳多与他同行，意在把这个年轻人训练成为一名商人。

布吉亚位于伊斯兰领土内，包括北非的一部分和西班牙南部三分之一领土，当时处于穆瓦希德人的什叶派王朝统治之下，其王朝位于远西的马拉喀什。因此，年少的列奥纳多在一个大穆斯林城里受到各方面学术的熏陶，估计接触过中世纪伊斯兰数学家如花拉子米和奥马·海亚姆等人的著作。不久，他父亲派他踏上商业旅途，足迹遍及地中海沿岸，包括埃及、叙利亚、西西里（直到1194年一直是诺曼王国，之后德国的霍亨斯道芬王朝继承了这个王国的统治权）、法国和拜占庭等。

应该有很多来自意大利贸易城市的其他年轻人加入类似的旅途。然而，列奥纳多天生就是一名数学家，在旅途中，他从拜占庭的希腊人和穆斯林人那里吸取了他们的知识精华，并且通过穆斯林人学习了很多波斯、印度和中国的知识。当他于1200年左右返回比萨并永久定居时，很有可能已经拥有了比当时西欧的任何人甚至是世上的任何人都广博的算术和代数以及其他学科的知识。

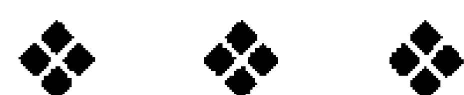


按当时的标准，《算盘书》极具创新的思想，并且产生了深远的影响，是随后的300年内一本最好的数学教科书。它介绍了包括0在内的阿拉伯（即印度的）数字，因而受到西方世界的广泛赞誉。事实上，这本书的开篇是这样写的：

> 一共有9个印度数字：9，8，7，6，5，4，3，2，1。使用这9个数字和阿拉伯语中称为zephirum的符号0可以表示任何数值，这正是本书要介绍的内容。

这本书共有15章，前7章的内容是使用这些“新”数字进行计算的入门知识，同时附有很多实例。剩余部分是算术、代数和几何的问题集，有的是关于商人和工匠的问题，还有的是如同前面我们已讲过的兔子数量等娱乐字谜问题，兔子数量的内容安排在第12章。

尽管《算盘书》在算术和历史方面都非常令人惊叹，但是它不是斐波那契的最有代数意义的著作，后来可能写于1225年的两本著作才充分显示了他的代数技能。我只选出第一本，因为它引出了本章将要涉及的主题：三次方程。



1225年前后或者就是在1225年，德国皇帝把宫廷设在比萨。这位皇帝就是弗雷德里克二世，他是那个时代最具魅力的人物，还有人称他是端坐于欧洲王位的第一位现代人。史蒂芬·伦西曼在其《十字军东征》的第三卷中详细描述了他的主要性格。弗雷德里克很享受被罗马教皇本人两次逐出教会的荣耀（人们不得不认为他确实很享受），这是当时盛行的教皇与皇帝之间的一种权力游戏。从那时起他就不喜欢天主教会。上文所提到的沃尔什教授在他用429页的篇幅所写的13世纪的赞美歌中，只提到了弗雷德里克一次，说他是那个世纪最聪明、最有教养的统治者。

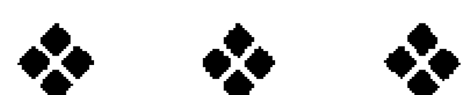
在那个年代，斐波那契的数学才能在比萨很有名，他还与弗雷德里克宫廷里的一些学者保持着良好的关系。所以，当弗雷德里克来到比萨时，斐波那契理所当然地成为了他的座上宾。

在弗雷德里克的宫廷里有一位叫巴勒莫·约翰尼斯的人，我没有找到太多关于这个人的介绍。有一份文献说他是一位马拉诺，也就是说是一位信奉基督教的西班牙犹太人。总之，约翰尼斯似乎对数学很了解，所以弗雷德里克要他给斐波那契出一些问题，来测试一下这个人的才能。

其中一个问题是求解三次方程$x^3+2x^2+10x=20$。斐波那契是否当场解决了这个问题我不知道。但是，他的确在他于1225年出版的著作中对这个问题给出了详细的描述。这本书有一个很长的拉丁文书名，通常把它

简称为《花》（*Flos*[6]）。这个方程的实根是1.368 808 107 821 372 6…。斐波那契给出的解非常接近这个数，只是在小数点后第11位错了。

我们不知道斐波那契是如何得到结果的，他没有告诉我们。也许他使用了几何方法，如奥马·海亚姆为了相同的目的而使用了相交曲线那样的几何方法。关于他对这个三次方程的处理值得关注的不是他的解是什么，而是他的分析过程。通过谨慎的推理，他首先指出这个解不是一个整数，然后又指出这个解也不是一个有理数。他还说这个解不是平方根，或者任意有理数和平方根的组合。这种对三次方程的分析是中世纪代数的杰作。该分析过程关注的是解的性质而不是解的实际值，可以说它已经预示了600年后在思索方程解的过程中将发生的一场伟大革命（后文会介绍这场革命）。



斐波那契没有用我们熟悉的十进制方法表示那个三次方程的解，而是类似于古巴比伦人那样用六十进制表示，如$1°22'7''42'''33^{\text{IV}}4^{\text{V}}40^{\text{VI}}$。这个表达方法的意思是

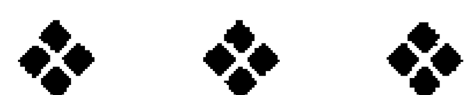
$$1+(22+(7+(42+(33+(4+40\div 60)\div 60)\div 60)\div 60)\div 60)\div 60$$

上式的计算结果是1.368 808 107 853 223 5。前面说过，这个结果到前10个小数位是正确的。事实上，中世纪后期代数学发展的巨大障碍是没有好的数值、未知量和算术运算的书写方法。在欧洲，斐波那契的阿拉伯数字得到普及是一个伟大的进步，但是在十进制数值符号被真正应用于小数部分之前，这第一次“数字革命”仍是不彻底的。

未知量和它的幂的表示情况更糟糕。抛开纯几何表示，正如我早前说过的那样，穆斯林代数学家都用文字说明一切，通常用阿拉伯语shai（事物）或jizr（根）来表示未知量，而用mal（财产或财富）表示未知量的平方，用kab（立方）表示未知量的立方，用组合形式表示高次幂，如用mal-mal-shai表示五次幂，等等。丢番图更快捷的记法知识都保存在君士坦丁堡[7]的希腊图书馆里，但是穆斯林和西方那些信仰基督教的数学

家似乎不了解这些，或者觉得没有了解的必要。

像斐波那契这样的早期的意大利代数学家效仿穆斯林人，把他们的文字翻译成拉丁语或意大利语：radix表示根，res、causa或cosa表示事物，census表示财产，cubus表示立方。到了14世纪末，后3个词分别被缩写成co、ce和cu，这是一个进步，是一位名叫卢卡·帕乔利的意大利人在1494年出版的一本著作中归纳的，这本著作通常被称为帕乔利的《算术大全》[8]。帕乔利的记法比丢番图的ς、Δ^{Y}和K^{Y}使用范围更为广泛，但是缺乏想象力。尽管《算术大全》中原创的东西很少，但是对商业计算来说使用起来很方便，因此一直深受欢迎。帕乔利被认为是复式簿记之父[9]。



我就这样漫不经心地跃过了269年，而没有介绍其间所发生的任何事情。一方面这是著作者的特权：我想先说说一般三次方程求解，然后引出卡尔达诺，这是本书中第一位有血有肉的人物。另一方面，斐波那契与帕乔利之间的200多年的确也没有什么值得一说的事情发生。

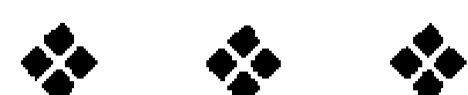
13世纪、14世纪和15世纪初确实有几位名副其实的代数学家。很多代数史中都列出了他们的一些贡献。例如，范德瓦尔登几乎用了近6页纸来介绍比萨的梅斯托洛·达尔迪，此人在14世纪中期处理了二次方程、三次方程和四次方程，把它们分成198个类型，并使用独创性的方法解决了某些特殊类型的方程。

尽管这位专家值得关注，但对一些二流人物却没什么需要了解的东西。只是，到了15世纪后半叶，印刷书籍的广泛传播加快了代数学的发展步伐。

这些进展并非只发生在意大利。法国人尼古拉斯·丘凯于1484年写了一份手稿（但直到1880年才得以出版），题目是《算术三编》。在这份手稿中，他引入了未知数的幂的上角标用法（当然这一用法不完全是我们现代的形式：他使用12^3代表$12x^3$），并把负数本身作为一种实体来处理。德国人约翰尼斯·威德曼1486年在德国的莱比锡讲授了第一堂代数课，

他也是在1489年出版的一本书中第一个使用现代加号和减号[10]的人。

除了丘凯这份很少受到关注的手稿外，这一方面的其他著作仍然使用中世纪末对未知量及其幂的记法，在法语中未知量是chose，在德语中是coss[11]。直到1540年一般三次方程的解发现之前，都没有什么非常有意义的重大发现。下面我就介绍1540年前后那段引人入胜的往事。



这段往事的中心人物是吉罗拉莫·卡尔达诺。他于1501年出生于帕维亚，1576年殁于罗马，一生中的大部分时间是在米兰或米兰附近度过的，因此他认为那里是他的故乡。

卡尔达诺是一位传奇式的人物，今天我们可以说他就是一部"作品"。他的传记就有好几部，第一部就是他自己写的《我的生平》，这本书是他在临终前写的。这本自传里有一个他的著作目录，长达几页纸。他总计有131本印刷出版的著作，还有111本未印刷出版的著作手稿，另外还有170份手稿因为自己不满意，他就直接销毁了。

他的很多著作都成为了欧洲的畅销书。例如，《安慰》这本书于1573年首次被翻译成英语，是一本劝慰悲伤的书，莎士比亚读过这本书。《哈姆雷特》中著名的"生存还是毁灭"的独白非常类似于《安慰》中关于睡眠的评述句子。据说当哈姆雷特在舞台上说出这句独白时，拿的就是这本书。

医学是卡尔达诺第一和主要的兴趣，也是他谋生的手段。他最早出版的著作也是医学著作，其内容主要是介绍一些药品常识，并嘲讽当时一些奇怪、肯定有害的医疗活动。（卡尔达诺说他用两周时间写了这本书。）他50岁时已成为继安德烈·维萨里之后欧洲第二知名的医生。当时的名流都争先恐后地找他看病。他似乎非常不喜欢旅行，一生中只有过一次长途跋涉：1552年去苏格兰治疗约翰·汉密尔顿的哮喘病。此人是苏格兰罗马天主教的最后一名教主。卡尔达诺的酬金是2 000金克朗。这次治疗似乎非常成功，因为汉密尔顿一直活到1571年，当时他已任满教

主的职位，因为参与暗杀苏格兰玛丽皇后的丈夫达恩利伯爵的活动而被吊死在斯特灵一个露天的绞刑架上。

在著作权法出现之前，写书甚至写畅销书也不是致富之路，顶多只是以间接方式宣传自己。卡尔达诺的第二收入来源是赌博和算命。在从苏格兰出发经伦敦回家的路上，卡尔达诺给年轻的爱德华六世（亨利八世的儿子）算卦，预言说他会长寿，只要躲过23岁、34岁和55岁时的病魔即可。不幸的是，爱德华在不到一年后就死了，年仅16岁。卡尔达诺还关注其他形式的预言。他甚至声称已经发明了一种东西——前额皱纹占卜法，根据脸部各种缺陷能读出人的性格和命运。在他关于这一话题的一本书中这么说道："妇女的左脸颊上酒窝左边一点若长有疣，她最终会被丈夫毒死。"

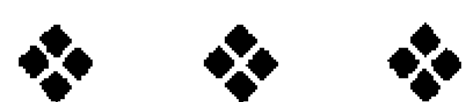
卡尔达诺对赌博的痴迷有可能到了成瘾的地步，而且有可能毁了他，但是事实上他是一个分析力很好的聪明棋手（在那个时代下国际象棋一般都是为了钱），而且是熟知数学概率的高手。他写了一本关于赌博的书《机遇博奕》[12]，其中就有对骰子和纸牌游戏的详细数学分析。

卡尔达诺在实践科学和理论科学方面很好地诠释了真正的文艺复兴精神。他的著作中含有大量打捞沉船和测量距离的设计、机械、仪器和方法的示意图。当罗马帝国皇帝查尔斯五世[13]于1548年来到米兰时，卡尔达诺因为为皇帝的马车设计了一个悬挂装置而在游行行列中获得了一个非常荣耀的位置。（查尔斯身患痛风病，旅行对他来说很痛苦。这对一个把欧洲领土从大西洋扩展到波罗的海的男人来说，真是一件不幸的事情。）在今天的法国和德国汽车上使用的传动轴仍然以卡尔达诺的名字命名。

卡尔达诺的人生低谷是他的儿子詹巴迪斯塔被判处死刑之时。他非常喜欢这个儿子并对其寄予厚望。这个男孩爱上一个不值得爱的女人并与她结了婚。这个女人生了三个孩子后，嘲讽詹巴迪斯塔说，这三个孩子都不是他的。于是他用砒霜将其毒死。（在16世纪，对于意大利妻子来说，被丈夫毒死似乎是一种职业危害。）很快他就被捕了，而且在执行死

刑之前受了严刑拷打。他还不到26岁。在卡尔达诺余生16个年头，这一可怕的事件一直折磨着他。然后，就在他的生命即将结束的时候，他自己也因是异教徒而被反改革的权威们抓捕。我们不知道给他定的是什么罪，他在自传中没有说这些，大概是被要求发誓保持沉默。在监狱被关押了一个月后，他被释放，但是不能离开家门，不允许再公开讲课或出版书籍。

尽管一生中经历了那么多的奇遇和不幸，但他还是安静地死在了床上，时间是1576年9月20日，享年75岁。这个时间与他几年前投掷骰子算出的时间完全吻合。有人说他给自己下毒或者把自己饿死，仅仅是为了保证这一时间与卜算吻合。按照他的性格这种推测也是完全有可能的。



卡尔达诺在代数历史中的伟大成就就是他的著作《大术，或论代数法则》。这本书是"一部伟大的艺术品，或者是关于代数法则的第一本书"，包含三次方程和四次方程的一般解，还在数学文献中第一次郑重其事地给出了复数。人们通常称其为《大衍术》，它于1545年在纽伦堡首次出版。

卢卡·帕乔利在《算术大全》中曾经列出两种没有可能解的三次方程：

$$(1)\ n = ax + bx^3$$

$$(2)\ n = ax^2 + bx^3$$

也即，"未知量与其立方之和等于一个数"和"未知量的平方与其立方之和等于一个数"。未被帕乔利列出的第三种没有可能解（我不知道他为什么没有列）的类型是"未知量加上一个数等于其立方"：

$$(3)\ ax + n = bx^3$$

对于我们来说，上面的第三种类型与第一种类型一样，但这是因为在处理过程中我们取了负数。在卡尔达诺的时代，人们刚刚知道负数的独立存在。

正如16世纪初的一些人指出的那样，一位名叫希皮奥内·德尔·费罗的人发现了第一种类型的三次方程的一般解。德尔·费罗是波洛尼亚

大学的数学教授，大约生活在1456年到1526年。我们不知道他解决第一种类型的三次方程的确切时间，也不知道他是否还解决了第二种类型的三次方程。他从没有发表他的解。

在德尔·费罗去世之前，他把“未知量加上其立方”的解的秘密告诉了他的学生，一个名叫安东尼奥·玛丽亚·菲奥利的威尼斯人。于是这个可怜的家伙就以中世纪数学家的身份出现在所有历史书中。我不怀疑历史学家们的判断，但是在如此重大的代数事件中，把他这个传播者误称为一名数学家，对菲奥利来说是一件不幸的事，这样他在数学上如此平庸之事实就如此这般在各个时代传播开来。总之，得到“未知量加上其立方”的秘密之后，他决定从中赚点钱。在当时文化活动非常活跃的意大利北部地区，这样做并不困难。那时学者们面临的境况是：很难得到赞助，大学职位的薪水又不太理想，而且也没有终生职位制度。为了谋生，他必须推销自己，例如，他可以与其他学者公开竞赛。对这样的竞赛投入的赌金越多，推销得就越好。

其中就有一位数学家在这样的公开竞赛中获得了声名，他就是尼科洛·塔塔里亚，威尼斯的一位教师。塔塔里亚来自于距威尼斯以西160公里的布雷西亚。当他13岁的时候，一只法国军队洗劫了布雷西亚，而且展开了屠杀。塔塔里亚虽然幸存了下来，但是他的下颌却受了严重的创伤，致使他口吃。他的姓塔塔里亚的意思就是“口吃的人”。当时那个时代就这样，一个人的姓氏由地位、父姓和昵称决定。塔塔里亚是有一定建树的数学家，他著有一本关于火炮技术中的数学的著作，又是把欧几里得《几何原本》译成意大利语的第一人。

1530年，塔塔里亚把关于三次方程的一些论述转交给布雷西亚的一位本地人，这个人叫朱安·托尼尼·柯伊，他在那个镇上教数学。在这些交换过程中，塔塔里亚说已经发现了第二种类型的三次方程的一般解法，不过他承认他还不能解决第一种类型的三次方程。

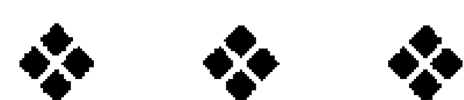
菲奥利这名数学才能平庸之人不知从哪里听说了这些交换和塔塔里

亚的声明。他可能认为塔塔里亚是诈骗，也可能确信他自己是知道如何求解第一种类型三次方程（这个秘密是德尔·费罗告诉他的）的唯一一个人，所以他向塔塔里亚发出了挑战。每一个人都要向对方出示30个问题，并在1535年2月22日将对手问题的30个答案送给公证人。输的一方要宴请赢家30次。

由于没有太重视菲奥利的数学才能，最初塔塔里亚没有对这次竞赛做充分的准备。但是有人传言说，菲奥利在10年前就从一位数学大师那里学习了如何求解“未知量加上其立方”。这时塔塔里亚才开始上心，他倾注了自己的全部才能去寻找第一种类型的三次方程的一般解。在2月13日上午，塔塔里亚仅用了几个小时就解开了这个问题。正如他推测的那样，菲奥利的所有问题都是关于第一种类型的三次方程的问题，这也就是菲奥利的全部数学才能了。

塔塔里亚的问题（我们只有其中前4个问题）似乎是第二种类型和第三种类型的混合型。有一点可以肯定，此时塔塔里亚已经掌握了只有一个实数解的任意类型的三次方程，也就是说那些有正判别式的三次方程。有负判别式的三次方程（因此有3个实根）可能通过处理复数而得到解决，当时还没有发现复数。

总之，塔塔里亚能够解决菲奥利的所有问题，而菲奥利不能解决塔塔里亚的任何问题。塔塔里亚赢得了荣誉，但放弃了赌金。卡尔达诺的自传[14]评论到：“与一个可怜的失败者面对面进餐的场景对他没有吸引力。”



卡尔达诺从达科伊那里听到了塔塔里亚取胜的消息，达科伊与塔塔里亚是同乡，并在1530年与塔塔里亚交流了关于三次方程的某些看法。与塔塔里亚交流之后，达科伊搬到了米兰。在意大利北部，数学教师很少，于是卡尔达诺聘请达科伊来教授他的部分学生。似乎正是从达科伊那里卡尔达诺得知了菲奥利与塔塔里亚之战的完整细节，以及塔塔里亚

与达科伊五年前的交流情况。此时，卡尔达诺正在写一本书，他预定的书名是《算术、几何和代数之实践》。也许他想如果塔塔里亚真得到了这个三次方程的解，那么把他的结果写入自己这本书里该多好。于是他开始着手一场游戏，想从塔塔里亚那里窃取这一秘密。

接下来这二个人之间的交往[15]读起来非常有趣。卡尔达诺就像一个熟练的钓鱼者正在垂钓一条鱼一样戏弄塔塔里亚，变换着方式戏弄他，从傲慢讽刺到甜蜜的诱惑。他们之间的书信交流时间为1539年的一月份至三月份。卡尔达诺投出的最吸引人的诱饵是把塔塔里亚介绍给意大利最有权力的男人之一阿方索·德阿瓦洛斯，他是整个伦巴底的地方长官（地位仅次于查尔斯五世皇帝）和驻扎在米兰附近的帝国军队的司令官。塔塔里亚关于火炮技术的著作在之前不久已经出版，卡尔达诺声称已经买了两本，一本是给自己，另外一本是给自己的朋友，就是这位长官。他的这位朋友长官非常想见见这本书的作者。

塔塔里亚急急忙忙来到米兰，并在卡尔达诺的家里住了几天。塔塔里亚的举动就如苍蝇直奔蜘蛛网一样，自投罗网。不幸的是，这位长官当时没在米兰。但卡尔达诺对待他的这位客人却真如皇室般尊重。最后，在3月25日，塔塔里亚答应把“未知量加上其立方”的秘密告诉热情款待他的主人，但他坚持要卡尔达诺发重誓不发表这一结果。卡尔达诺郑重其事地发誓，塔塔里亚就以25行诗的形式写下了这个三次方程的解。这首诗是这样开始的：

Quando che'l cubo con le cose appresso
Se agguaglia a qualche numero discreto...
（“当未知量的立方和未知量总和等于某个整数时，……”）

按照塔塔里亚的自述，他离开卡尔达诺家之后就为这次举动而后悔了。他回到威尼斯的家中开始反思。卡尔达诺写信给塔塔里亚询问这首诗中某些地方的解释，但是塔塔里亚的回应非常粗暴。卡尔达诺的算术

书在5月份出版，其中并没有介绍三次方程的解，此时塔塔里亚的愤怒才得以平息。但是，那年的夏天，他听说卡尔达诺又开始写另外一本专门研究代数的书。他们之间的交锋演变成了这种局面，一边是塔塔里亚对卡尔达诺的怀疑和愤怒，一边是卡尔达诺对塔塔里亚的安抚，这种局面一直持续到了1540年。

《大衍术》出版于1545年。从1540年初两人之间的最后一次书信到1545年《大衍术》出版之间的这5年是代数历史中的重要时期。卡尔达诺在得到未知量和其立方之和的秘密之后，进一步给出了三次方程的一般解。

研究了不可约的情况后，他开始意识到三次方程总是存在三个解。为了研究这种情况，他必须使用复数的概念。卡尔达诺的确带有很多疑问和迷茫，对此我们不应感到惊讶。当时人们甚至觉得负数都有些神秘。虚数和复数就显得更加不可思议了。（对今天的大多数人来说也是这样的。）

下面是《大衍术》第37章中的内容，研究的是四次方程而不是三次方程：把10分成乘积等于40的两部分。

> 不要考虑得太多，把（$5+\sqrt{-15}$）乘以（$5-\sqrt{-15}$），得到[$25-(-15)$]，后面一项实际上就是+15。因此，其结果是40。……这实在是太奇妙了……

的确很奇妙。卡尔达诺肯定是仔细考虑了很长时间，艰难地做出了这样的突破。他的思想也偏入到其他方向。他找到了求得近似解的数值方法，并逐渐形成了解与系数之间关系模式的概念，因此瞥见了直到150年后数学家才开始触及的领域。

卡尔达诺有一位助手。这要回溯到1536年，那年他雇用了一个名叫费拉里的十四岁小孩当仆人。他发现这个孩子非常聪明，已经能够读写，所以他就把这个孩子提升为自己的私人秘书。费拉里通过校对1540年的算术一书的手稿学会了数学。我们可以认为在卡尔达诺与三次方程搏斗

的过程中，他一定与这位年轻的秘书谈论过这个问题。

我们这样认为的一个原因就是，在1540年费拉里找到了一般四次方程的解，正如我在数学知识中所提到的那样，这其中包括求解三次方程。因此，只有先会解三次方程，费拉里才能出版四次方程的解。三次方程一定是他从卡尔达诺那里学会解的，而卡尔达诺却已经向塔塔里亚发誓绝不会发表这个结果。

另外，在1526年德尔·费罗去世到1535年菲奥利与塔塔里亚之战开始的这几年里，早就有传闻说菲奥利从已故的德尔·费罗那里得到了未知量与其立方之和的解。为了逃避道义上这种进退两难的局面，1543年，卡尔达诺和他的秘书费拉里来到博洛尼亚旅行，准备与菲奥利这位德尔·费罗的继承人谈谈，此时他已成为德尔·费罗的女婿和论文管理人。仔细翻阅这些论文之后，卡尔达诺和费拉里了解到塔塔里亚不是解决未知量与其之和立方的第一人。这一事实为他们提供了道义上的突破口，卡尔达诺出手了，他在《大衍术》里给出了三次方程和四次方程的全部解。他认定德尔·费罗是发现未知量与其立方之和的解的第一人，而塔塔里亚只是再次发现而已。

塔塔里亚当然大发雷霆，此前整整五年时间他一直在静静地翻译欧几里得和阿基米德著作。随后的骂战持续了三年，只是卡尔达诺没有参战，只留下费拉里顶替他参战。直到1548年8月10日，费拉里与塔塔里亚在米兰又进行了一次学术性竞赛，这场骂战才结束。我们只有一份简报，疑为塔塔里亚的笔述，从中可看出是塔塔里亚输了这场竞赛。

塔塔里亚死于1557年，他是带着愤怒和痛苦离开人世的。他从没有发表过自己的三次方程解的成果，在其论文中也没有找到翻译的版本。毫无疑问他独立解决了未知量与其立方之和的问题，但是这一荣誉却被首先建立三次方程分类的德尔·费罗和掌握了三次方程全部解且成为四次方程求解方法教父的卡尔达诺平分了。

注　解

[1] 为了证明这一点，我们称这个斐波那契数列的第n项为u_n。所以u_1是1，u_2也是1，u_3是2，u_4是3，等等。现在，利用这些斐波那契数为系数构建下面的多项式：

$$S = x^{n-1} + x^{n-2} + 2x^{n-3} + 3x^{n-4} + 5x^{n-5} + 8x^{n-6} + \cdots u_{n-2}x^2 + u_{n-1}x + u_n$$

用x乘以上面等式的两边，得到xS。重复这一过程得到x^2S。从x^2S中减去S和xS。根据斐波那契数列的性质，你会看到右边的很多项都消掉了。例如，项x^{n-6}的系数是(21−13−8)等于零。最后剩余部分是：

$$\left(x^2 - x - 1\right)S = x^{n+1} - u_n x - u_{n-1}x - u_n$$

让x分别等于$x^2-x-1=0$的两个平方根，这样就消掉了S，从而得到一个关于两个未知量u_n和u_{n-1}的联立方程，消掉u_{n-1}，得到待证结果。

[2] 二项式定理给出了$(a+b)^N$的展开公式。在N=4的特殊情况下，可得$(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4$，这个展开式就是我在这里使用的公式。

[3] 根据DSB中关于斐波那契的介绍，这本书的书名拼写是Liber abbaci，而不是通常所写的Liber abaci，供喜欢争论的读者参考。英译文为*The Book of computation*（《计算书》），而不是*The Book of the Abacus*（《算盘书》）。在意大利语中，书的名字同歌剧的名字一样，不必每个词都大写。

[4] 也就是说斐波那契出生于著名的比萨斜塔开建前的一两年内。这一斜塔始建于1173年，耗时180年。才建到第三层时，它的倾斜就已经很明显了。

[5] 现在这个城市是阿尔及利亚的贝贾亚市，距阿尔及尔以东120英里。

[6] Flos是花的意大利文，含义是“非常好的著作”。

[7] 学术政治家迈克尔·普塞路斯曾经在1051年~1075年服务于拜占庭帝国，他是迈克尔七世（1071年~1078年在位）的重臣。他肯定知道丢番图的文字符号体系。

[8] 该书的全名是*Summa de arithmetica, geometria, proportioni et proportionalita*,《算术、几何、比例及比例性的概括》。顺便提一下，现在我们已经进入了在欧洲印刷书籍的时代。在威尼斯印刷的帕乔利的著作是首批在欧洲印刷的数学著作。

[9] 帕乔利后来的一本著作因达芬奇为其插图而赢得了相当高的声望。

[10] “加”和“减”的意大利文分别为*piu*和 *meno*；后来就像未知量的幂的表示一样，这些符号都被简写了，一般写成*p*和*m*。

[11] 事实上，15世纪和16世纪的德国代数学家被称为Cossists，而代数被称

为the Cossick art。1577年英国数学家罗伯特·雷科德出版了一本书，书名是《砺智石；算术的第二部分，包括取根及方程法则的代数应用》。这是使用现代等号的最早的印刷著作。

[12] 卡尔达诺的这本书在他去世后才出版。该书的英文版作为附录放在了奥斯丁·欧尔为卡尔达诺写的传记中。

[13] 查尔斯五世是威尔第的歌剧"唐·卡洛"中的幽灵僧侣的原型。圣罗马皇帝的头衔是选举产生的。为了顺利得到它，查尔斯花费了一百万达克特去贿赂选民。他是由罗马教皇加冠（1530年2月在波洛尼亚）的最后一位皇帝。他那个时代的大部分人都认为他是西班牙国王（第一个使用查尔斯这个名字的罗马国王，因此有时会与西班牙的查尔斯一世混淆），实际上他在佛兰德斯登基，他的西班牙语很蹩脚。

[14] 指奥斯丁·欧尔为卡尔达诺写的传记《卡尔达诺，赌博学者》（1953年出版）。顺便提一下，欧尔为卡尔达诺写的传记是我看过的传记当中可读性最好的长篇记述，但是不幸的是早已绝版了。如果想要了解卡尔达诺的占星术可以参见安东尼·格拉夫顿所著的《卡尔达诺的宇宙》（1999年出版）。关于卡尔达诺的书很多，仅传记就至少有三本。

[15] 在欧尔的书中对他们二人之间的交往过程有详细的记载。对于他们之间发生的事情，欧尔整整用了55页来描述，在正文中我已经把这一记述浓缩成几段文字。原文值得读者看看。对此还有另外一个全面的记载，但是事件和时间与欧尔的略有不同（我使用的是欧尔的版本），请参见马丁·诺加德发表于《国家数学杂志》第13期上的"卡尔达诺—塔塔里亚之战拾零"一文。该文已被美国数学协会在2004年年志《巴比伦的福尔摩斯》中再印。

第5章 放飞想象

现代欧洲初期在代数学上有两个伟大的进步：一是得到了一般三次方程和四次方程的解，二是发明了现代符号体系，即用字母表示数的体系。这里所说的“欧洲初期”，指的是从君士坦丁堡的失陷（1453年）到威斯特伐利亚的和平（1648年）这两个世纪期间。

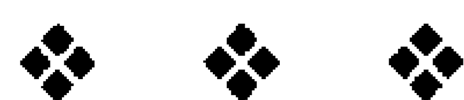
第一个进步是由意大利北部的数学家在大约1520年到1540年间完成的，而卡尔达诺与费拉里在1539年到1540年的合作研究时期是其中最富有创造力的时期。我已经在第4章中讲述了他们的故事。

第二个进步主要是两名法国人的研究成果：弗朗索瓦·韦达[1]（1540—1603）和勒奈·笛卡儿（1596—1650）。与之齐头并进的另一个成果就是复数的缓慢发现，及其逐渐被人们接受，成为了标准的数学工具。而后者的发现严格说来更多是属于算术的范畴（是关于数值的），而不是代数范畴（主要研究的是多项式和方程）。当然，我们已经介绍过，复数的发现从代数中获得了启发。如果绘制一个“不可约”三次多项式的图形（参见图CQ-6），显然它有3个实零点；然而如果利用代数公式求解相应的方程，且不允许使用复数，那么就没有实数解！

说复数和代数关系密切的依据还有不少。例如，它为非常重要的代数概念“线性无关”提供了最初的线索，我随后将讨论这一概念，而且这一概念将引出向量和张量理论，使现代物理学成为现实。如果把3和5相加，就会得到8：3的性态与5的性态相融，它们融合起来形成8，就如

同两滴水珠的融合。然而，把3与5i相加，得到一个复数3+5i，这就如同油滴与水滴一样不融合，这就是线性无关。

因此我说有必要讲述复数的发现，并且这段历史很有意思。正如我们在前面看到的那样，带着相当认真的态度处理这些奇怪生物的第一位数学家是卡尔达诺。第一位大胆地向其发出挑战的数学家则是拉斐尔·邦别利。



邦别利来自波洛尼亚，希皮奥内·德尔·费罗在那里教书。他出生于1526年，德尔·费罗去世的那一年。因此他正好是卡尔达诺下一代的人。历史就是这样，一代人要拼命去理解的东西可能对下一代人来说容易得多。在邦别利大约19岁时，《大衍术》出版，他正处于接受这一著作影响的最佳年龄。

邦别利是一位土建工程师。他接手的第一桩大项目是土地开垦工作，其具体任务就是排干意大利中部佩鲁贾附近的沼泽地。这一项目从1549年开始一直持续到1560年，并且获得了巨大的成功，从而让邦别利在职场中一举成名。

邦别利非常欣赏《大衍术》，但是觉得卡尔达诺的解释还不是十分清晰。就在他20岁的时候，他有了这样的野心，想写一本自己的代数著作，一本使初学者完全能够掌握这门学科的书。这本书的书名是《代数》，出版于1572年，是在邦别利死前几个月出版的，所以他在这本著作上投入了大约25年的时间，从20岁出头到40好几。这本著作显然经过了多次改动和改版，最终出版的版本是作者深思熟虑的结果。

大约在1560年左右，邦别利在罗马遇到了在那里当数学老师的安东尼奥·玛丽亚·帕齐，并与他一起讨论数学。帕齐说他曾在梵蒂冈图书馆发现了一份关于算术和代数的手稿，是古希腊“某个叫丢番图”的作者所写的。这两个人仔细查阅了这份手稿，然后决定翻译它。这一翻译最终没有完成，但是毫无疑问，邦别利在研究丢番图著作的过程中受到

了很多启发。他在《代数》一书中描述了143个丢番图的问题。正是由于邦别利所做的工作，那个时期的欧洲数学家才首次了解到丢番图的工作。

回想一下，尽管丢番图没有把负数当作一个数学对象的概念，而且也不认为它们是问题的解，但他还是允许它们在中间计算过程中像影子一样存在，并为此构建了符号规则。卡尔达诺几乎也用同样的方式处理复数，虽然认为复数本身没有意义，但它却是得到实数问题的实数解的很有用的工具。

邦别利对负数和复数的处理方法更加成熟。他全面接受了负数，比丢番图更加清楚地重新陈述了符号法则：

più via più fa più
meno via più fa meno
più via meno fa meno
meno via meno fa più

这里più的意思是“正”，meno的意思是“负”，via的意思是“乘”，fa的意思是“得到”。翻译过来就是：正正得正，负正得负，正负得负，负负得正。

在《代数》中，邦别利着手处理不可约三次方程，寻找方程$x^3=15x+4$的解。他使用了卡尔达诺的方法，得到

$$x=\sqrt[3]{2+\sqrt{-121}}+\sqrt[3]{2-\sqrt{-121}}$$

他利用一些算术技巧计算出两个立方根分别是$2+\sqrt{-1}$和$2-\sqrt{-1}$。把这两个解加起来就得到$x=4$。（另一对解是$\left(-2-\sqrt{3}\right)$和$\left(-2+\sqrt{3}\right)$，但是邦别利没得出来。）

这里所讲的复数的地位如同丢番图的负数一样，只不过是从一个“实数”问题得到一个“实数”解的过程中的内部技巧而已。可以说它们起到了一种催化作用。实际上邦别利认为它们是“诡辩的”。他认可它们为一种合法的工具，甚至还给出了它们相乘时的“符号法则”：

più di meno via più di meno fa meno
più di meno via meno di meno fa più

meno di meno via più di meno fa più
meno di meni via meno di meno fa meno

这段话的意思是：从负到正乘以从负到正得负，从负到正乘以从负到负得正，从负到负乘以从负到正得正，从负到负乘以从负到负得负。这里的“从负到正”指$+\sqrt{-N}$，而从负到负指$-\sqrt{-N}$，其中N是某个正数。via和fa分别是“乘”和“得到”的意思。所以上面诗文的第三行的意思是：如果把$-\sqrt{-N}$与$+\sqrt{-N}$相乘，那么将得到一个正的结果。这一结论是对的，其结果是N。通常的符号法则（负乘以正）得出负的结果，把平方根平方得到的结果是$-N$，$-N$的负数是正数。

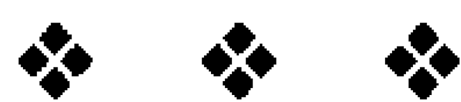
邦别利的《代数》在对数学的理解上向前迈出了一大步，但是他还是受到了没有好的符号体系的影响。对于下面的公式：

$$\sqrt[3]{2+\sqrt{-3}}\times\sqrt[3]{2-\sqrt{-3}}$$

他写成：

Moltiplichisi, R.c.[2 *più di meno* R.q.3] per
R.c.[2 *meno di meno* R.q.3]

这里的R.q.的意思是平方根，R.c.的意思是立方根。注意，他还使用了中括号。符号体系发展了，对卡尔达诺那一代人的表述有了改进，但是改进并不是太多。



对于法国人来说，16世纪不是一个幸福的时代。弗朗西斯（1515—1547）统治的大部分时间和这个国家的大部分财富都消耗在与皇帝查尔斯五世的战争中。1559年《卡托—康布雷齐和约》缔结，战争告一段落。但这个劳动力消耗殆尽的国家喘息未定，法国天主教徒和后来常被称为雨格诺教徒[2]的新教徒就彼此开始了大屠杀。

直到《南特敕令》（1598年）颁布后他们才停止屠杀，或者说在某种程度上中断了87年。之前的36年间[3]，法国人之间发生了8次内战和一次朝代更迭（波旁王朝取代了瓦卢瓦王朝）。这些战争并非纯粹的宗教之战。

区域情结、社会等级以及国际政治等诸多因素都是这些战争的起因。西班牙国王菲利普二世，这个世上最大的麻烦制造者，竭尽全力把局势搅乱。至于说社会阶层，雨格诺教徒在城市中产阶级当中势力很强大，并且估计有一半的贵族也是新教徒。而在大部分地区的几乎所有农民仍然是天主教徒。

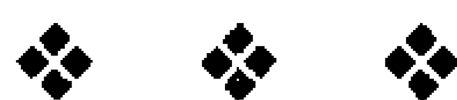
韦达于1540年出生在一个雨格诺教徒的家庭里，父亲是一名律师。他于1560年从波瓦第尔大学毕业，获得法律学位。他毕业不到两年，法国宗教战争就开始了，其标志事件是香槟地区瓦塞小镇的雨格诺教徒遭到屠杀。

战争决定了韦达后来的人生。他放弃了律师行业，成为一个贵族的家庭教师。随后于1570年搬到巴黎，显然是希望得到政府的雇用。当时年轻的查尔斯四世在位，但是他的母亲凯瑟琳·梅迪奇（她还是西班牙国王菲利普二世的岳母）才是真正的掌权者。为了保持王权强大且不受制于各种势力，她挑拨雨格诺教徒与天主教之间的关系。这个政策决定了整个16世纪60年代到80年代法国的历史进程，而且经常产生荒谬的结果。韦达在巴黎的时候正是查尔斯批准于圣巴多罗谬之夜（1572年8月23日）开始屠杀雨格诺教徒的时候。然而第二年，韦达这个雨格诺教徒却被国王任命为布列塔尼地区的政府官员。

查尔斯死于1574年，凯瑟琳的第三个儿子亨利三世继位。6年后韦达返回巴黎担任这位皇帝的顾问。然而，凯瑟琳最小的儿子死于1584年，留下一个没有继承人的瓦卢瓦王朝。虽然亨利三世已经结婚而且只有33岁，但是众所周知他是一个同性恋，经常穿着女人的衣服，对宫廷大事也不上心。很难想象他会生出一个儿子来。因此他的远房亲戚波旁皇族的纳瓦拉·亨利就成为了王位的合法继承人。然而，这个亨利是一个新教徒，这使得法国内外的天主教感到恐慌。宫廷混战变得更加激烈。韦达被排挤了出去，被迫在布尔纽夫海湾波伏娃滨海的一个小镇里的家中休了5年的长假。从1584年到1589年的这5年时间是韦达数学创造的鼎盛

时期，当时他近50岁，能在数学上有所创造真是不同寻常。当时法国的宫廷政治斗争错综复杂，给写数学史的历史学家留下种种疑团，他们都不知道应该埋怨谁。

就在韦达返回宫廷的4个月后，亨利三世被暗杀，居然是如厕时被一剑刺中要害。纳瓦拉·亨利成为亨利四世，这是波旁王朝的第一位皇帝。这位新皇帝是一名新教徒，这对韦达很有利，他很高兴成为新皇帝的随从。然而，即使天主教不得不同意让一名竞争对手成为王位的候选人，他们也不会让这位亨利四世顺利上任。西班牙国王菲利普非常喜欢自己的女儿，并为她的利益同法国宫廷派系秘密私通。这些秘密联络都是用密码写在信上的。亨利发现身边有韦达这位数学家，便委派他破解西班牙人的密码。经过几个月的努力之后，韦达破解了这些密码。当菲利普得知他自认为不可破解的密码已被破解时，他对罗马教皇抱怨说亨利使用了魔法。

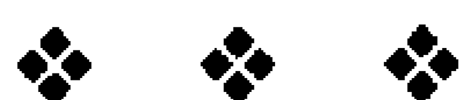


韦达一直为亨利四世服务，直到1602年的12月他被宫廷免职为止。他返回家乡一年后就去世了。

除了成功破解密码，韦达在为宫廷服务期间在数学上取得的辉煌战绩发生在1593年。那一年，说荷兰语的数学家艾德里安·范·罗门出版了一本名为《理想的数学》的书，此书介绍了当时所有的杰出数学家。荷兰驻亨利四世宫廷大使对亨利说这本书中没有一个法国人。为了证明这一点，他给这位皇帝看了罗门书里的一个问题，这位作者为这个问题的解设立了一个奖项。这个问题是寻找一个数，使其满足一个45次方程，$x^{45}-45x^{43}+945x^{41}-12300x^{39}\cdots$。显然，这位外交官（他也太不圆滑了）是在嘲讽法国数学家不能解决这个问题。亨利派人请来韦达。韦达当场就找到了一个解，并且在第二天就给出了另外22个解。

当然，韦达知道罗门不是随意就给出一个方程，它一定是罗门自己知道如何求解的方程。处于那个时代，韦达有着丰富的三角学知识，而

在当时三角学[4]正是发展迅猛的一个数学分支。他的前两本书就是三角学表格集。三角学是研究圆的弧长和弦长之间的数量关系的科学，全是用正弦、余弦以及它们的幂的公式表示的。韦达根据这个方程中前几项的系数快速心算后判断出他看到的是这样的一个式子：在设$x=2\sin\alpha$时关于$2\sin 45\alpha$的多项式。于是，三角学帮他求出了解。（至少帮他给出了23个正数解。还有22个负数解，韦达忽略了它们，显然是认为负数没有意义。）



韦达40多岁时在海边的5年流放生活的成果，就是写成了一本名为《分析术入门》（*Introduction to the Analytic Art*）的书。该书代表代数学向前迈进了一步，同时也表明向后退了一小步。向前的一步是第一次系统地使用字母来表示数值。这一思想的萌芽阶段要追溯到丢番图，但是韦达是有效地分配字母并制定了不同的量、使用不同范围的字母的第一人。这就是现代符号体系的开始。

韦达字母符号体系不同于以往所有的设计方案，它不仅仅局限于未知量。他把量分成两大类：一类是未知量，或者说是“要求的量”（寻求的目标）；还有一类是已知量，或者说是“已给定的事项”（数据）。他用大写元音字母A、E、I、O、U、Y表示未知量，而用大写的辅音字母B、C、D等表示给定量。例如，方程$bx^2+dx=z$用韦达字母体系表示为：

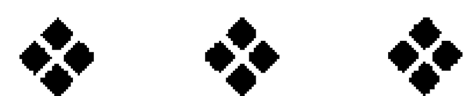
B in *A* Quadratum, plus *D* plano in *A*, aequari *Z* solido.

他的*A*是未知量，就是我们的*x*。其他符号都是数据。

他的方程中的plano和solido就是我上面所提到的后退。韦达深受古代几何的影响，希望他的代数能够建立在严格的几何概念之上。他的这种认识就迫使他遵循同次性法则，也就是说方程中的每一项必须有相同的维度。除非另外指定，否则每一个符号代表适当长度的一条线段。在上面给定的方程中，*b*和*x*（韦达的*B*和*A*）都是一维的。于是bx^2就是三维。因此dx必须也是三维的，同样*z*也是。因为*x*是一维的线段，所以*d*必须是

二维的，因此是D plano。类似地，z必须是三维的：Z solido。

你能明白韦达的观点，但是同次性法则限制了他的风格，同时也使其代数的某些部分难以理解。这多少也有点奇怪，一个能够如此巧妙地处理45次多项式的人居然被古典几何和它的三维牢牢束缚了。



与邦别利相比，韦达对方程的处理在某种程度上还不够“现代”。如上文所述，他反对负数，不能接受负数作为解。他对复数的态度甚至更落后。他只是在一本关于几何的书中处理过三次方程，在那里他根据用$\sin\alpha$表示$\sin 3\alpha$的公式提出了一种三角学解法。

然而，韦达仍是研究方程的先驱，他点燃的蜡烛在200年后成为了一个巨大的灯塔。在他的有生之年，这一特殊的发现没有发表。在他去世12年后，他的苏格兰朋友亚历山大·安德森发表了他的两篇论文。在名为《论方程的完整化》的第二篇论文中，韦达为伽罗华理论、群理论和所有现代代数的诞生开辟了一条研究通道。

考虑二次方程$x^2+px+q=0$。假设方程的两个解，也就是使这个方程成立的两个数是α和β。如果x是α，或者x是β，而没有其他情况，那么下面的等式一定成立：

$$(x-\alpha)(x-\beta)=0$$

因为α和β是使得这一方程成立的x的所有值，所以上面的式子一定是原始方程的一个重写形式。现在，如果用通常的方法把括号乘开，那么这个重写方程为

$$x^2-(\alpha+\beta)x+\alpha\beta=0$$

把上面这个方程与原来的方程相比较，下面的式子一定成立：

$$\alpha+\beta=-p$$
$$\alpha\beta=q$$

这里，我们得到了方程解与系数之间的关系式。也可以用同样的方法处理三次方程$x^3+px^2+qx+r=0$。如果这个方程的解是α、β和γ，那么有

$$\alpha+\beta+\gamma=-p$$
$$\beta\gamma+\gamma\alpha+\alpha\beta=q$$
$$\alpha\beta\gamma=-r$$

对于四次方程$x^4+px^3+qx^2+rx+s=0$，同样有

$$\alpha+\beta+\gamma+\delta=-p$$
$$\alpha\beta+\beta\gamma+\gamma\delta+\alpha\gamma+\beta\gamma+\alpha\delta=q$$
$$\beta\gamma\delta+\gamma\delta\alpha+\delta\alpha\beta+\alpha\beta\gamma=-r$$
$$\alpha\beta\gamma\delta=s$$

对于五次方程$x^5+px^4+qx^3+rx^2+sx+t=0$，同样有

$$\alpha+\beta+\gamma+\delta+\varepsilon=-p$$
$$\alpha\beta+\beta\gamma+\gamma\delta+\delta\varepsilon+\varepsilon\alpha+\alpha\gamma+\beta\delta+\gamma\varepsilon+\delta\alpha+\varepsilon\beta=q$$
$$\gamma\delta\varepsilon+\alpha\delta\varepsilon+\alpha\beta\varepsilon+\alpha\beta\gamma+\beta\gamma\delta+\beta\delta\varepsilon+\alpha\gamma\varepsilon+\alpha\beta\delta+\beta\gamma\varepsilon+\alpha\gamma\delta=-r$$
$$\beta\gamma\delta\varepsilon+\gamma\delta\varepsilon\alpha+\delta\varepsilon\alpha\beta+\varepsilon\alpha\beta\gamma+\alpha\beta\gamma\delta=s$$
$$\alpha\beta\gamma\delta\varepsilon=-t$$

这些等式应该这样描述：

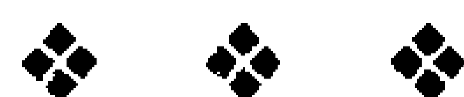
所有解之和$=-p$

所有解两两相乘再求和$=q$

所有解三三相乘再求和$=-r$

依此类推

一个未知量的五次以下方程的这些解是由韦达首先写出来的。韦达下一代的法国数学家吉拉德在他的《代数中的新发现》一书中把这些关系一般化到任意次方程。《代数中的新发现》一书出版于1629年，是在安德森发表韦达论文的14年之后。艾萨克·牛顿爵士继承并发扬了他们的衣钵……我还是先把这段故事讲完吧。



借用主持人的口吻：笛卡儿无需介绍。他当过兵，从过政（他在战场上幸存下来，却没逃脱政治阴谋的魔爪），他是哲学家、数学家，是前三代波旁皇帝脚下的一个法国臣民。笛卡儿成年后，社会也一直动荡不安，三十年战争爆发，英国内战战火燃起，朝圣风行。那也是法国红衣

主教黎塞留当政的时代，是瑞典国王古斯塔夫·阿道尔夫节节取胜的时代，是米尔顿和伽利略名扬天下的时代。笛卡儿是法兰西的民族英雄，尽管他更喜欢生活在荷兰。他的出生地当时叫拉哈耶，而法国大革命之后为了纪念他，被重新命名为笛卡儿（波瓦第尔以北50公里的地方）。同56年前的韦达一样，笛卡儿在波瓦弟尔大学获得律师学位后步入社会。

笛卡儿因两件事而闻名于世：一是写了名言“我思故我在”，二是后来以他的名字命名的坐标系统，即用数值表示平面上所有点的坐标系统。在笛卡儿几何中，决定一个点的数值是那个点到两个固定坐标轴的垂直距离，而两条坐标轴则是平面内相互成直角的两条线。东西距离通常称为x，南北距离通常称为y。点的笛卡儿坐标见图5-1。

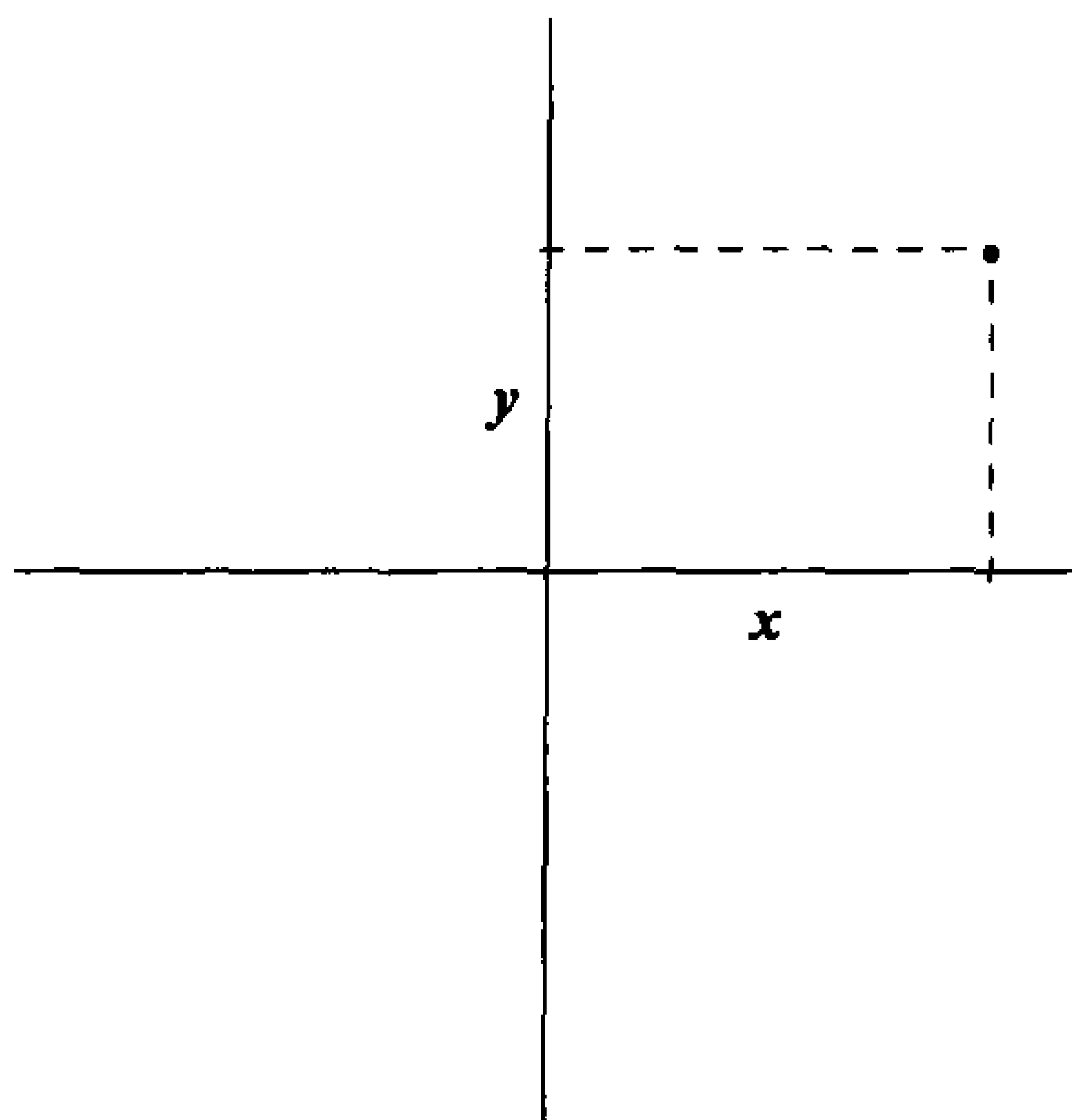


图5-1 笛卡儿坐标

“我思故我在”这句话确实是笛卡儿写的，但是他给出的坐标系并不是现在这个这么精确的笛卡儿坐标系。他关于坐标系的基本想法都写在了他的著作《几何学》（1637年出版）中，但是他使用的坐标轴不是彼此成直角的。

不过，《几何学》中所包含的这些思想已足以引发几何和代数的革命，使几何学代数化。回想一下韦达的同次性法则，它就是建立在数值

是几何对象的思维映像这一思想之上的。甚至当韦达遇到45次幂时，他的大脑里也立即反映出一个几何概念，把一个圆弧分成45等份。笛卡儿看得更远，他认为几何对象正是数值的一个方便的表现形式。笛卡儿还给出了具体的例子：两条线段长度的积不必一定表示为一个矩形的面积，还可以把它表示成另外一条线段。

就表示方法本身来说这不是原创思想，但是笛卡儿进一步在其上建立了他的几何框架，砍断了连接代数与古典几何的最后绳索，使他的“分析几何”冲出重围飞入宇宙。笛卡儿采用了更清晰更有序的代数记法，所以做到了这一点。他从16世纪德国代数学家那里吸取了加号、减号和他们的平方根符号（他还在原来的平方根符号上面加了一个横线，把√变成$\sqrt{\ }$）。他将幂用上标形式表示，不过平方没用上标形式，还是写成aa而不是a^2，直到19世纪末有些数学家才开始写成a^2。

笛卡儿最重要的贡献大概在于，给我们带来了现代文字符号体系。在这一体系下，字母表中前几个小写字母表示已知数（数据），字母表中后几个小写字母表示要求的量（寻找的目标）。阿尔特·约翰逊在他的著作《古典数学》中关于代数符号讲述了下面这样的故事。

> 使用字母x来表示未知量源于一个巧合的事件。在打印《几何学》的过程中，打字员遇到一件进退两难的事情。在文本排版的时候，打字员发现字母表的最后几个字母不够用了。他问笛卡儿，书中诸多方程中使用字母x、y还是z是否很重要。笛卡儿回答说，用这三个字母中的哪个字母表示未知量都行。于是这个打字员就选择x表示大多数未知量，因为字母y和z在法语中的使用频率要高于x。

事实上，阅读《几何学》时你会觉得是在看一本现代数学课本。我想这本书是最早的一本现代数学读物。唯一古怪的事情是书中没有我们现代人所使用的等号：笛卡儿使用一个类似于左端被切掉了的无穷符号

的小符号作为等号。

引入一个好用的文字符号表示是数学史上的一个重大进步。这样做的不只是笛卡儿，韦达也用字母表示数值的系统。他们表示未知量的灵感则来源于丢番图。

在这里不提一下英国人托马斯·哈里奥特也许是不公平的。哈里奥特的生活年代是1560年到1621年，是韦达和笛卡儿之间的一代人。他为沃尔特·雷利爵士①工作过许多年，至少陪同雷利爵士去维吉尼亚探险过。他是一位了不起的数学家。很有可能受到了韦达的启发，他使用字母表中的字母表示数据和未知量。由于通晓方程理论，哈里奥特考虑了负数解和复数解。遗憾的是，直到他死后[5]许多年，这些成果才得见天日，他活着的时候没有发表数学方面的作品。数学历史学家都喜欢争论笛卡儿到底从哈里奥特那里借鉴了多少东西，《几何学》出版6年前，哈里奥特的代数著作就发表了（版式很糟糕）。然而，据我所知，尚没有任何确凿证据可以证明笛卡儿借鉴了哈里奥特的成果。

总之，首次使得文字符号系统广为人知且方便易用的是笛卡儿。他的符号系统相当完善，在随后的4个世纪中都无需做本质上的改动即可应用。这一体系不仅使数学家受益，而且还激发了莱布尼茨为人类思想创造符号论的梦想，以至于所有关于真或假的争论都可以通过计算解决。莱布尼茨说这样的一个体系将"放飞想象力"。当我们把笛卡儿的数学描述与之前的代数学家的描述相比较时，可以看到，一种好的符号体系的确能够放飞想象力，把复杂而高层次的思维过程转换成容易掌握的符号运算。

1649年，古斯塔夫·阿道尔夫的女儿瑞典王后克里斯蒂娜邀请笛卡儿教她哲学。她派出一艘船把笛卡儿接到斯德哥尔摩，与法国大使居住

① 沃尔特·雷利（约1552—1618），英国航海黄金时期的一位船长，提出了"征服海洋即可征服世界"的观点，深得当时英国女王伊丽莎白一世青睐。——编者注

在一起。这位王后习惯早起，而笛卡儿却从童年起就养成了睡到11点才起床的习惯。1649年到1650年那个冬天极其寒冷，笛卡儿不得不每天早晨5点就在寒风呼啸中赶到宫殿教王后哲学。可怜的笛卡儿得了肺炎，1650年2月11日去世，享年54岁。艾萨克·牛顿当时只有7岁。

注 解

[1] 说英语的数学家们一般把韦达（Viète）的名字读作Vee-et，母语是英语且比较固执的人更喜欢用这个跟著名蔬菜汁饮品一样的名字。有时候这个名字也被写成拉丁语的形式Vieta。

[2] 通常是这样称呼的，但不完全准确。雨格诺教徒是加尔文教徒，法国新教徒不全是。也许有很多人不高兴被称作雨格诺教徒。然而这个名字一直沿用至今，是因为在这一类书籍中经常被引用，“雨格诺教徒”可以认为是“早期法国新教徒”的同义词。这个词的词源不太好理解。

[3] 尽管这一时期英国与法国一度有着良好的外交关系，然而在莎士比亚的《亨利四世II》中有很多针对法国人的笑话和攻击。

[4] 巴斯德在《科学传记辞典》中说“韦达的所有数学研究都与天文学和宇宙学的工作相关”。天文学和三角学的结合起源于对天体的研究，源自于对星星高度的计算和预测等。

[5] 哈里奥特临死时相貌非常恐怖。他得了鼻癌，这可能是因为他在维吉尼亚开始养成的烟瘾造成的恶果。在生命的最后8年里，他的脸逐渐地被侵蚀掉。

第二部分

泛　算　术

第6章

狮子的爪子

从16世纪末到18世纪初，尽管经历了内战（1642－1651）、军事独裁（1651－1660）、光荣革命（1688年）及两个朝代的更替（1603年斯图亚特王朝推翻都铎王朝，1714年汉诺威王朝推翻斯图亚特王朝），不列颠群岛上还是出现了几位相当优秀的数学家。

我已经提到过哈里奥特，他精巧的文字符号体系并没有受到广泛的关注（也许笛卡儿关注过）。苏格兰人约翰·纳皮尔虽然不是一位知名的代数学家，但是他发现了对数并于1614年将此公布于世。他还普及了小数点。威廉·奥特雷德，英国的一位乡村牧师，写了一本关于代数和三角学的著作，并发明了乘号。约翰·沃利斯是第一个使用笛卡儿的分析几何技术和记法的人，他是早已不在人世的哈里奥特的拥护者，认为笛卡儿从哈里奥特那里得到了这一记法。

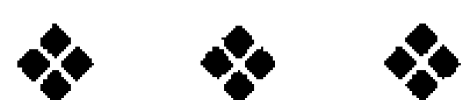
然而，所有这些人物都只是牛顿登场的前奏。这位杰出的天才被公认为是科学史中最伟大的人。他于1642年的圣诞节出生[1]，是林肯郡一个比较富裕的农场主的遗腹子。介绍他的生活历程和性格特点的作品已经很多了。下面是我自己以前写过的一段话。

> 牛顿的生活经历并不丰富。他从没有走出英格兰东部，也没有过从商经历和从军历史。尽管当时英国制宪史中发生了一些重要事件，但是他似乎对公共事件不感兴趣。代表剑桥大学

> 成为英国议员并没有给他带来在政治舞台上展现的机会。牛顿也没有过什么亲密爱人。据他自述，到死他都是一个处男，这一点似乎勿庸置疑。他同样淡薄友谊，对于出版作品也常常是迫于无奈，因此常常用假名，因为害怕“能够得到公众的关注并维系它也许增加了我的知名度，但这种事情更要命的是使我的研究水平下降”。他行事小心谨慎、不冷不热、心不在焉，所以与同事之间总是为一些小事而争吵，这导致他与人交往时从来不能心情愉快、畅所欲言。正如这位英国人对自己的评论一样，他是一个“冷酷无情的人”。[2]

我忍不住要讲一个我特别喜欢的牛顿故事，尽管我知道这个故事已广为人知。1696年，瑞士数学家约翰·伯努利给欧洲数学家提出了两个难题。牛顿看到这两道题的当天就解决了它们，并把答案交给伦敦皇家科学院院长。这位院长把答案寄给伯努利，但是没有告知是谁做的。伯努利一看到这个匿名答案就知道是牛顿做的，他说：“看到他的爪子，我们就知道这只狮子。”

这只强大的狮子在代数历史上刻上了重要的一笔。



牛顿[3]因对科学的贡献和发明微积分而闻名，但是人们却不知道他是一名代数学家。事实上，1673年到1683年间他在剑桥大学开过代数讲座，这所大学的图书馆里还存放着他的课程讲稿。很多年后，当他离开学术界去当皇家造币厂厂长时，他的剑桥接班人威廉·惠斯顿把这些讲稿集结成书发表了，书名是《泛算术》。牛顿非常不情愿地同意了出版此书，他似乎从来就没有喜欢过这本书。他拒绝署名，甚至打算把这本书都买下来销毁了。牛顿的名字既没有出现在1720年出版的英语版本上，也没有出现在1722年出版的拉丁语版本上。[4]

引起代数历史学家兴趣的不是《泛算术》本身，而是年轻的牛顿于

1665年或1666年所写下的一些简短的笔记，在他的第一本著作《数学集锦》中可以找到这些笔记。它们是用英文而不是拉丁文写的。笔记是这样开始的：

> 每一个诸如$x^8+px^7+qx^6+rx^5+sx^4+tx^3+vxx+yx+z=0$的方程的根的个数都等于其维数，所有根之和是$-p$，每两个根之和是$+q$，每三个根之和是$-r$，每四个根之和是$+s$，依此类推。

这些注释没有陈述任何定理。但是，它们暗示了一个定理。这个定理太令人震撼了，数学家们（实际上还有《数学集锦》的编辑们）就把这个暗示的定理称作牛顿定理。

在给出这个定理之前，我需要解释对称多项式的概念。为了便于处理，我要考虑3个未知量，分别称它们是α、β和γ。下面是若干含有这3个未知量的对称多项式：

$$\alpha\beta+\beta\gamma+\gamma\alpha$$
$$\alpha^2\beta\gamma+\alpha\beta^2\gamma+\alpha\beta\gamma^2$$
$$5\alpha^3+5\beta^3+5\gamma^3-15\alpha\beta\gamma$$

下面是α、β和γ的非对称多项式：

$$\alpha\beta+2\beta\gamma+3\gamma\alpha$$
$$\alpha\beta^2-\alpha^2\beta+\beta\gamma^2-\beta\gamma^2+\gamma\alpha^2-\gamma^2\alpha$$
$$\alpha^3-\beta^3-\gamma^3+2\alpha\beta\gamma$$

第一组多项式与第二组多项式之间有什么区别呢？表面看来，它们之间的差异好像是：在第一组，发生在α身上的事情，也同样发生在β和γ身上；发生在β身上的事情，也同样发生α和γ身上；发生γ身上的事情，也同样发生在α和β身上。所谓的事情就是加、乘和各种组合，它们都以均等的方式发生在所有未知量上。而第二组多项式就没有这些性质。

这已经说得很清楚了，但是，对于对称多项式的条件，用数学语言精确描述是：如果以任意方式置换α、β和γ，都将得到相同的表达式。

实际上，有5种置换α、β和γ的方式。

- 交换β和γ，留下α不变。
- 交换α和γ，留下β不变。
- 交换α和β，留下γ不变。
- 用β取代α，γ取代β，α取代γ。
- 用γ取代α，α取代β，β取代γ。

（注意：也许某个数学家告诉你会有6种置换方式，第六种是“恒等置换”，即什么都不做。下一章我将就此阐述我自己的观点。）

如果你对第一组多项式中的任一个多项式做上述任意置换，最后得到的多项式都与原来的一样，当然需要重新排列一次。例如，如果对$\alpha\beta+\beta\gamma+\gamma\alpha$做最后一种置换，最后得到的是$\gamma\alpha+\alpha\beta+\beta\gamma$，这与原来的多项式一样，只是写法不同。

还有另外一种检测方法，当一个多项式很大且很难处理时，一个有用（尽管不总是正确）的方法是给α、β和γ赋任意值，于是就能得出一个结果。如果你以所有可能的顺序把这些值赋给α、β和γ时结果都一样，那么这个多项式就是对称的。如果我用6种可能的方式给α、β和γ分别赋值0.550 34、0.812 17和0.161 10，计算得出$\alpha\beta^2-\alpha^2\beta+\beta\gamma^2-\beta^2\gamma+\gamma\alpha^2-\gamma^2\alpha$的6个对应值，3个值为0.066 353 6，3个值为−0.066 353 6。这就不是一个对称多项式：置换未知量得到两个不同的值。（这件事本身很有趣，为什么是两个值？我稍后再解释。）

这些原则可以扩展到任意多个未知量和任意复杂程度的表达式上。下面是一个包含两个未知量的11次对称多项式：

$$\alpha^8\beta^3+\alpha^3\beta^8-12\alpha-12\beta$$

再看一个包含11个未知量的二次对称多项式：

$$\alpha^2+\beta^2+\gamma^2+\delta^2+\varepsilon^2+\xi^2+\eta^2+\theta^2+\iota^2+\kappa^2+\lambda^2$$

并不是所有对称多项式都同等重要。一些多项式称为*初等对称多项式*。3个未知量的初等对称多项式是：

1阶：$\alpha+\beta+\gamma$

2阶：$\beta\gamma+\gamma\alpha+\alpha\beta$

3阶：$\alpha\beta\gamma$

我上面给出的对称多项式的例子中，第一个是初等的，其他两个不是。

对于任意数目的未知量，有如下初等对称多项式。

1阶：所有未知量加起来（“所有单项”）。

2阶：所有可能的未知量对加起来（“所有对”）。

3阶：所有可能的三元组加起来（“所有三元组”）。

依此类推。

如果处理的是有n个未知量的多项式，那么这个列表有n行，因为n个变量不可能制造出$(n+1)$元组。

现在我可以给出牛顿定理了。

牛 顿 定 理

任意n变量对称多项式可以用n变量初等对称多项式表示。

所以，尽管我上面给出的两个对称多项式不是初等对称多项式，但是它们可以用我刚才给出的初等对称多项式表示。第二个对称多项式写起来很容易：

$$\alpha^2\beta\gamma+\alpha\beta^2\gamma+\alpha\beta\gamma^2=\alpha\beta\gamma(\alpha+\beta+\gamma)$$

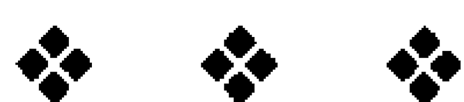
第三个对称多项式写起来有点麻烦，但是很容易就可以验证：

$$5\alpha^3+5\beta^3+5\gamma^3-15\alpha\beta\gamma=5(\alpha+\beta+\gamma)^3-15(\alpha+\beta+\gamma)(\beta\gamma+\gamma\alpha+\alpha\beta)$$

按照习惯，同时为了便于理解，我们通常研究固定数目个未知量的多项式（上面的是3个未知量的多项式），初等对称多项式通常用小写的希腊字母表示，用下标表示阶。对于上面的情况，3个未知量的1阶、2阶和3阶初等对称多项式分别记为σ_1、σ_2和σ_3，所以我可以把最后的等式写成：

$$5\alpha^3+5\beta^3+5\gamma^3-15\alpha\beta\gamma=5\sigma_1^3-15\sigma_1\sigma_2$$

这就是牛顿定理：任意数目个未知量的任意对称多项式可以用初等对称多项式σ表示。



这一切与解方程有什么关系呢？为什么没有关系呢，回头看一下第5章中的那些α、β和γ的多项式，这些多项式是韦达处理时使用的。它们是初等对称多项式！对于一个五次方程$x^5+px^4+qx^3+rx^2+sx+t=0$，如果它的解是$\alpha$、$\beta$、$\gamma$、$\delta$和$\varepsilon$，那么就有$\sigma_1=-p$，$\sigma_2=q$，$\sigma_3=-r$，$\sigma_4=s$，$\sigma_5=-t$，其中这些$\sigma$是5个变量的初等对称多项式，这些多项式我已经在第5章中写出来了。对于x的任意次多项式，类似的等式也成立。

正如我提到的那样，这些笔记让我们知道了牛顿定理，它们写于1665年或1666年，是在牛顿早期的数学生涯中写下的。当时他21岁，刚刚获得文理学士。由于突然爆发瘟疫，剑桥被迫停课，牛顿回到了农村的家里。两年后学校复课，为了获得学院奖学金和硕士学位，牛顿回到了学校。在农村的那两年时间里，牛顿建立了一些基本思想，这成为他后来在数学和科学领域诸多发现的基础。人们常说数学家在30岁之后就不做原创性的工作了，这种说法并不严格，但一般情况下的确如此，他们的思维模式和吸引他们兴趣的课题都是在他们早期的论著中找到的。

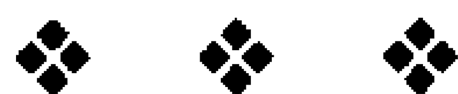
实际上，在做这些笔记的时候，牛顿头脑中有一个特殊的问题，这个问题是确定什么时候两个三次方程有一个共同解。然而，这个工作是建立在以下基础上的：

(1) 一般的对称多项式；

(2) 方程的系数与这个方程的解表示的对称多项式之间的关系。

这个问题是方程理论进一步发展的关键，无论是方程理论本身还是建筑在其上的一个全新的代数领域，所有的结果都由它而来。17世纪末，在突破三次方程和四次方程问题的120年后去给五次方程求解之际，诸如对称及用系数表示的多项式的解都是解决多项式方程理论中这个常见问

题的关键。



总体说来，与17世纪和19世纪相比，18世纪是代数发展较慢的时期。牛顿和莱布尼茨的微积分发明为数学发展开发了广阔领域，但是这不包括本书中我所指的代数，而是我们今天所说的“分析”领域，研究内容是极限、无穷序列、级数、函数、微分和积分等，在那个年代，分析是一门具有魅力的新领域，数学家们对它投入了极大的热情。

还有一门更具普遍性的数学学科被唤醒了，这就是韦达为代数而发展起来的现代符号学，而笛卡儿又凭借“放飞想象力”使数学研究更加容易。另外，对复数的进一步接纳也拓展了数学想象的边界。棣莫弗定理可以当作18世纪初期纯数学的代表，这一定理的完整版本最早出现于1722年，如下所示：

$$(\cos\theta + \mathrm{i}\sin\theta)^n = \cos n\theta + \mathrm{i}\sin n\theta$$

这一定理在三角和分析之间搭建了一座桥梁，也促使复数成为分析的必需之物。

上面讲的只是纯数学。随着科学的不断发展、工业革命的首次兴起以及宗教战争之后欧洲国家体制的逐步确立，君主和军队对数学家的需求也在增加。欧拉为弗雷德里克的无忧宫设计管道工程，傅里叶是拿破仑征服埃及时的科学顾问。

达朗贝尔在18世纪中期所做的微分方程的先驱性工作是非常有代表性的，拉普拉斯方程$\nabla^2\phi = 0$描述了一个量（密度、温度或电势）在一个平面或空间上均匀但非匀等分布的众多物理系统，可以看成是18世纪末应用数学的代表。

在某种程度上，代数是所有这些迷人发展的旁观者。一般三次方程和四次方程已经破解，但是还没有人考虑出在这个方向上如何进一步发展。韦达、牛顿和其他一两个最具想象力的数学家已经开始注意到多项式方程解的奇特对称性，但是如何从这些观察中获得数学结果还是毫无

头绪。

然而，还有一个问题，18世纪的数学家们为此奋斗了整整一个世纪，所以我应该在这里提出来。这个问题是寻找所谓的代数基本定理的证明，下文我称该定理为FTA（Fundamental Theorem of Algebra，代数基本定理）。我之所以用了“所谓”一词，是因为人们总是这样称呼这个定理，但是这个名称的含义还有待商榷。甚至有些数学家会用伏尔泰嘲讽神圣罗马帝国的口吻说，FTA不是什么基本也不是什么定理，更不属于代数的范畴。我希望马上澄清这一切。

FTA的内容很简单，如果用多项式方程稍微粗略地表示就是：*每一个方程都有解*。更精确的陈述如下。

代数基本定理

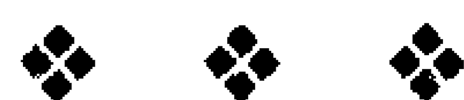
$x^n+px^{n-1}+qx^{n-2}+\cdots=0$是一个未知数$x$的多项式方程，这个多项式的系数$p$、$q$等是复数，$n$大于零，存在某个复数使这个多项式方程成立。

这里，普通的实数被理解为复数的特殊情况，实数a可以看成是a+0i。所以，我之前所给出的所有实系数方程都属于FTA的范畴。每一个这样的方程都有一个解，当然这个解可能是复数，如方程$x^2+1=0$，复数i满足这个方程（而且复数−i也满足这个方程）。

FTA是由笛卡儿最早在《几何学》（1637年）中陈述的，当时他是以一种假设的形式讲述的，因为他对复数还不熟悉。所有18世纪伟大的数学家都尝试着证明它。1702年莱布尼茨认为他已经驳斥了这个定理，但是在他的陈述中有一个错误，这个错误在40年后被欧拉指了出来。1799年伟大的高斯以它作为博士论文的主题。然而，直到1816年才有一个完整的证明，也是高斯给出的。

为了澄清FTA的数学地位，需要仔细研究一下证明。这个证明不难，只要熟悉复平面（参见图NP-4）并能够找到一本好的高等代数教科书[5]

就可以。下面是最粗略的证明。



代数基本定理的证明

复数同实数的情况一样，方程中幂较高的项可以很轻松地“淹没”幂较低的项，我曾在第3章后“数学知识：三次方程和四次方程”中提到此事。立方变大的速度快于平方，四次幂比立方更快，依此类推。（注意：对于复数的情况，单词“大”的意思是“远离原点”或等价于“有较大的模”。）因此，对于较大值的x，定理中的多项式方程看起来很像x^n，而其他项只起到微调的作用。

如果x是零，除了最后一项“常数项”外，这个多项式中的每一项都等于零。因此，对于较小的x值，这个多项式接近于它的常数项。（例如在$x^2+7x-12$中，常数项是-12。）

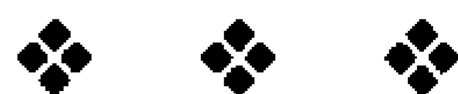
如果平稳而均匀地改变x，那么x^2、x^3、x^4以及所有更高次幂也都平稳而均匀地改变，只是它们改变的速度不同。它们不会突然从一个值跳到另一个值。

有了前面这三个事实之后，考虑所有给定较大模M的复数x。如果在复平面上标出它们，这些数会形成以M为半径的完美的圆。这个多项式的对应值形成一个半径为M^n的一个大得多的圆，但只是近似的。（如果一个复数有模M，那么它的平方有模M^2，依此类推。这一结果很容易证明。）这是因为x^n吞没了这个多项式的所有幂更低的项。

逐渐、平稳地将M缩小到零。所有模为M的复数所形成的完美圆也相应地缩小到原点。这个多项式的相应值也缩小，就像一个拉紧绳子的绳圈，从一个较大的以原点为圆心的近似圆，退化成这个多项式中常数项的那个复数。在这样的退化过程中，这一拉紧的多项式绳圈一定在某点经过原点。否则这些点怎么能收拢成那个复数呢？

这就证明了这个定理！那个缩小的绳圈上的点是对应于某个复数x的这个多项式的值。如果这个绳圈经过原点，那么这个多项式对于某个x

等于零。证毕。（你也可以花一点时间考虑一下这个多项式的常数项等于零的情况。）



从代数的角度看，不幸的事情是这个证明取决于连续这一事实。我要说的是，当x逐渐且缓慢地变化时，这个多项式的值也相应逐渐且缓慢地变化。这是成立的，但是它的成立仅仅是由于复数的性质，在复数系统中，你可以没有跳跃或颠簸地从一个数平稳滑动到另一个数，跨过中间无数个稠密排列的数。

并不是所有数值系统都如此慷慨。在现代代数中有各种各样的数值系统，我们可以在所有这些系统中建立多项式。像复数这样友好的系统并不多，FTA并不能在所有系统中成立。

因此，从现代代数的观点看，FTA是关于复数系统性质的描述，用现代的行话说这一性质称为代数封闭性。复数系统是代数封闭的，也就是说，任意一个未知量的以该数值系统的值为系数的多项式方程在该系统内都有一个解。FTA不是关于多项式、方程或数值系统的一般陈述。这就是为什么有些数学家会带着傲慢口吻告诉你它不是基本的。也许它可能是一个定理，但是它不是一个真正的代数定理，而是一个分析定理，连续概念属于分析的范畴。[6]

注　解

[1] 牛顿的出生日期是旧年历日期，1752年这一年历已被废除。如果根据我们现在使用的年历，他的出生日是1643年1月4日。这就是为什么他的出生日期有两个版本的原因。

[2] 这是我给帕特丽夏·法拉的书《牛顿：天才的制造》写的书评。我说牛顿对“公共事务不感兴趣”，指的是他那个时代的重大政治事件。从1696年开始，他任皇家造币厂厂长，他在这个岗位上尽职尽责，废寝忘食，而且很有创造力。他还活跃于皇家学会，从1703年开始一直到他去世，每一年他都蝉联会长。在反抗国王詹姆士二世的宗派威胁时，他勇敢地站出来为他的大学说话。虽然如此，我还是认为牛顿没有认真哪怕是短暂地考虑过政治、民族或国际事物。

[3] 如果你喜欢，也可以称他为“艾萨克爵士”。1705年他被安妮皇后封为爵士，是第一位获此殊荣的科学家。在1705年之前，严格地说，都应该称他为“牛顿”，1705年以后的他才能被称为“艾萨克爵士”。没有人会如此拘泥于这些细节，我也不想成为这样的先例。

[4] 无论是在拉丁文原版还是在英文翻译版都没找到。然而这本书的内容包含在《艾萨克·牛顿数学论文》的卷2之中。

[5] 我特别推荐1991年出版的迈克尔·阿廷的教科书《代数》（pp.527-530），此书的讲解清晰明了。

[6] 正如高斯以及后来的克罗内克所指出的，代数基本定理涉及一些深奥的哲学问题。如果想了解全面的讨论，请看哈罗德·爱德华的著作《伽罗华理论》。

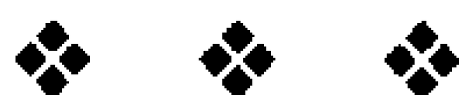
数学知识

单位根

在介绍一般三次方程求解时，我提到了1的立方根。它一共有3个立方根。1本身是1的立方根，因为1×1×1=1。1的其他两个立方根是：

$$\frac{-1+i\sqrt{3}}{2} \text{ 和 } \frac{-1-i\sqrt{3}}{2}$$

这两个根通常被称为ω和ω^2。如果对其中任何一个取立方，同时记住$i^2=-1$，就会发现答案总是1。另外，第二个数是第一个数的平方，第一个数是第二个数的平方。当然ω^2是ω的平方。稍微计算一下可知，ω是ω^2的平方（因为ω^2的平方是ω^4，而$\omega^4=\omega^3\times\omega$，而根据定义$\omega^3=1$）。

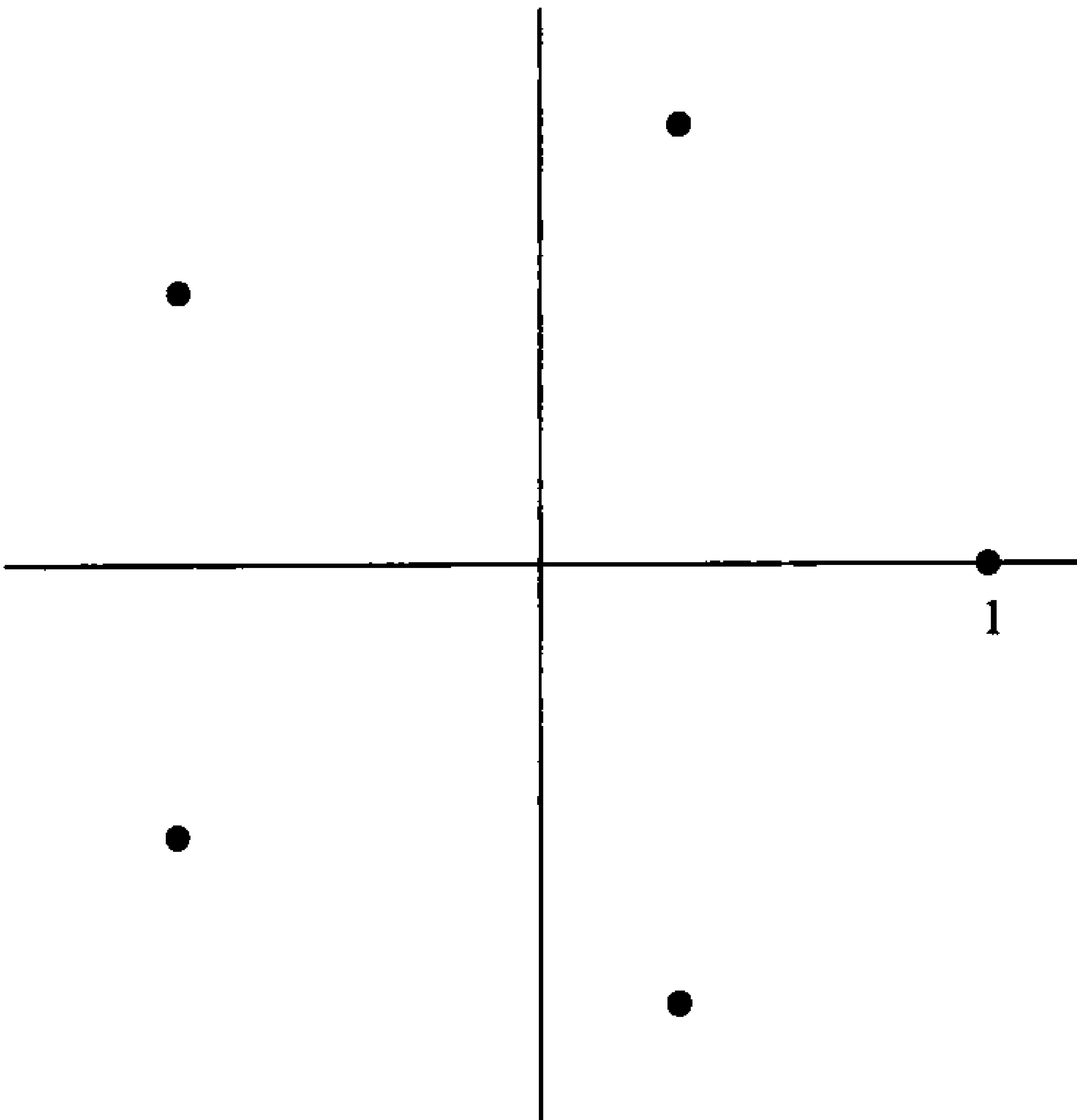


研究我们通常称为“单位”的1的n次根非常有趣，而且触及了几个不同的数学领域，其中包括古典几何和数论。但只有当数学家们能够很自如地处理复数时，才能真正展开这项研究，也就是说在18世纪中期才开始研究。伟大的瑞士数学家欧拉于1751年在一篇题为“论根和无理数的开方”的论文中对这个问题进行了广泛的探讨。

当然1的平方根是1和−1。1的立方根是1和前面我给出的两个数ω和ω^2。1的四次方根是1、−1、i、−i。这些根的4次幂都是1。欧拉证明了单位1的五次方根是：

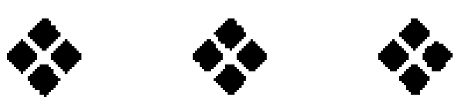
$$1,\ \frac{\left(-1+\sqrt{5}\right)+i\sqrt{10+2\sqrt{5}}}{4},\ \frac{\left(-1-\sqrt{5}\right)+i\sqrt{10-2\sqrt{5}}}{4},$$
$$\frac{\left(-1-\sqrt{5}\right)-i\sqrt{10-2\sqrt{5}}}{4},\ \frac{\left(-1+\sqrt{5}\right)-i\sqrt{10+2\sqrt{5}}}{4}$$

用数值表示就是：1、0.309 017+0.951 057i、−0.809 017+0.587 785i、−0.809 017−0.587 785i和0.309 017−0.951 057i。如果在复平面内标出这些数值，把实数部分标在东西向，而把虚数部分标在南北向，就有点像图RU-1的样子。



图RU-1　单位的五次方根

事实上，它们是以原点为中心的正五边形的顶点。换一种描述：它们分别位于以1为半径的单位圆的圆周上，而且它们把这个圆周平均分成五等份。表示“平分一个圆”的希腊语是cyclotomic。这些复数或者说是复平面上的点被称为cyclotomic points（分圆点）[1]。



这些数都是从哪里来的呢？我们是如何知道1的复数立方根是上面

给出的ω和ω^2的呢？动手解方程就知道了。

如果x是1的立方根，那么当然有$x^3=1$。换一种方式表示：$x^3-1=0$。而这恰好是一个可以求解的三次方程。事实上，因为我们知道$x=1$一定是其中的一个解，所以立即就能将其因式分解为这样的形式：

$$(x-1)(x^2+x+1)=0$$

所以，为了得到另外两个解，我们必须解这个二次方程。根据一般二次根公式可知，这个方程的解就是我所说的ω和ω^2（参见第1章注解[6]）。

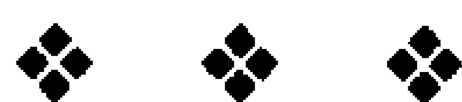
事实上，对于任意一个解为单位的n次方根的方程来说，方程$x^n-1=0$都可因式分解成如下形式：

$$(x-1)(x^{n-1}+x^{n-2}+x^{n-3}+\cdots+x+1)=0$$

而且除了1本身之外的其他单位根都可通过解下面这个方程得到：

$$x^{n-1}+x^{n-2}+\cdots+x+1=0$$

18世纪和19世纪的数学家花了大量笔墨来介绍上面这个方程的n个一般值。高斯在他的经典著作《算术研究》中用了整整一章的篇幅，英文译本中这一章占了54页。有时候我们把这个方程称为“割圆方程”，但现代数学家所说的“割圆方程”有更严格的意义。



高斯的最高成就是证明了正十七边形可以用古典的方式（即只用圆规直尺）构造出来。

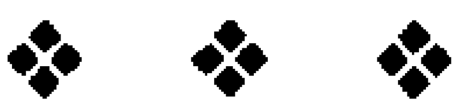
用我的话说就是，当且仅当在复平面内构成正多边形顶点的割圆点都可仅用整数和平方根符号写出来时，才可以这样构造正多边形。如前面我给出的$n=5$时的单位的五次方根就属于这种情况。所以正五边形可以用圆规和直尺构造出来。高斯证明了正十七边形也是如此。事实上，他写出了单位的十七次方根的实数部分：

$$\frac{-1+\sqrt{17}+\sqrt{34-2\sqrt{17}}+2\sqrt{17+3\sqrt{17}-\sqrt{34-2\sqrt{17}}-2\sqrt{34+2\sqrt{17}}}}{16}$$

虽然嵌套了三层，但都是整数和平方根，因此可以用直尺和圆规构

造。这是年轻的高斯的第一个伟大数学成就。这个成就太了不起了，所以人们把正十七边形刻在了他的出生地——德国不伦瑞克的纪念碑上。

高斯还证明，对于形如$2^{2^{k+1}}$的素数（而不能错误地说成是这种形式的任意数），上面的事实也是成立的。当k=0、1、2、3和4时，这些数分别是3、5、17、257和65 537，它们全是素数。然而，当k=5时，$2^{2^{k+1}}$是4 294 967 297，“正如伟大的欧拉第一个指出的那样”（我引用了高斯的说法），这个数不是素数。



单位根有很多有趣的性质。它不仅与古典几何关系密切，而且与研究素数、因数和余数的数论有着密切的联系。

为了初步了解单位根，要考虑单位的六次方根。它们是1、$-\omega^2$、ω、-1、ω^2、$-\omega$，而ω和ω^2是我们熟悉的单位的立方根。（我希望到目前为止大家应该很熟悉它们了。）如果你依次取这六个根中的每一个，然后再求它的一次、二次、三次、四次、五次和六次幂，就会得到下表中的结果。这些根被放置于表头标有α的最左栏。每一个根的各次幂按列排在一行。（任何数的一次幂就是这个数本身。这六个根的六次幂都等于1。）

α	α^1	α^2	α^3	α^4	α^5	α^6
1	1	1	1	1	1	1
$-\omega^2$	$-\omega^2$	ω	-1	ω^2	$-\omega$	1
ω	ω	ω^2	1	ω	ω^2	1
-1	-1	1	-1	1	-1	1
ω^2	ω^2	ω	1	ω^2	ω	1
$-\omega$	$-\omega^2$	ω^2	-1	ω	$-\omega^2$	1

在这个过程中，六个根中只有两个根$-\omega^2$和$-\omega$生成了所有六个根。其他根只能生成六个根的某个子集。这一结果与直觉一致，因为只有$-\omega^2$和$-\omega$是单位的六次方根，而其他的都是单位的平方根（例如-1）或单位的立方根（例如ω^2和ω）。

在n=6时，像$-\omega^2$和$-\omega$这样的单位的n次方根的各次幂可以生成所有的n次方根，我们称它们是单位的n次本原根[2]。“第一个”n次方根（按顺时针方向绕着复数图示的单位圆行进）总是本原根。其他的n次本原根是第k个，其中k是与n没有公因子的数。例如，单位的九次本原根应该是第一个、第二个、第四个、第五个、第七个和第八个九次方根。如果n是素数，那么除了1之外，单位的所有n次方根都是n次本原根。可见，现在我们说的素数和因子都属于数论的范畴。

注 解

[1] 似乎是1879年西尔维斯特第一次在这样的情况下使用了“分圆”一词。

[2] 不要把“单位的n次本原根”与数论中的术语“一个素数的本原根”相混淆。一个数g是一个素数p的本原根，如果用p除以g、g^2、g^3、g^4、……、g^{p-1}之后，余数是1、2、3、……、$p-1$的某个重排列。例如，8是11的本原根。如果取8的从1到10次幂，会得到8、64、512、4 096、32 768、262 144、2 097 152、16 777 216、134 217 728和1 073 741 824。这些数除以11后取余数，得到8、9、6、4、10、3、2、5、7和1。所以8是11的本原根。另外，3不是11的本原根。3的前10次幂是3、9、27、81、243、729、2 187、6 561、19 683和59 049。把这些数除以11后取余数，得到3、9、5、4、1、3、9、5、4和1。所以3不是11的本原根。事实上，（素数的）本原根的概念与我正文中的这个概念相关，但是不相同。因为11是一个素数，单位的每一个11次方根是单位的一个11次本原根，但是在数论的意义下11的本原根只是2、6、7和8。

顺便提一下，现在我可以解释术语“分圆方程”的“更严格的意义”。它是一个方程，其解全是单位的n次本原根。所以在n=6时，它应该是$(x+\omega)(x+\omega^2)=0$这样的方程，它的解是$x=-\omega$，$x=-\omega^2$。这个方程展开后是$x^2-x+1=0$。

第7章

攻克五次方程

我已经介绍了16世纪前半叶意大利数学家是如何发现一般三次方程和四次方程的解的。下一个挑战显然是五次方程：

$$x^5+px^4+qx^3+rx^2+sx+t=0$$

让我提醒一下诸位我们这里是要寻求什么。对于任意一个特定的五次方程，使用10世纪和11世纪数学家所熟悉的技术仅能够找到一个达到所希望精度的数值解。我们不知道的是代数解，一个如下形式的方程解：

$$x=[用p、q、r、s、t表示的某个代数表达式]$$

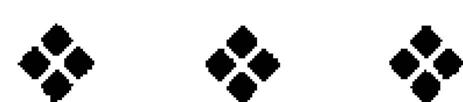
其中括号里的“代数”一词的意思是“只包含加法、减法、乘法、除法和开方（平方根、三次方根、四次方根、五次方根等）”。为了防止出现疏漏，我还是指定括号里的表达式只包含有限次运算。像以前那样给出四次方程一般解的表达式，此处给出五位方程的一般解正是人们梦寐以求的目标。

我们现在知道这样的解不存在。据我所知，第一位相信这一情况，也就是相信一般五次方程没有代数解的人是一位意大利人——保罗·拉菲尼（Paolo Ruffini）。他大概是在18世纪末（可能是1798年）确认了这一想法。第二年他发表了一份证明。（高斯在1799年所写的博士论文中也陈述了相同的观点，但是没有给出证明。）接着，拉菲尼分别在1803年、1808年和1813年发表了第二个、第三个和第四个证明。这些证明没有一个让当时的数学家感到满意，他们甚至没有关注这些证明。公认的真正

给出一般五次方程不存在代数解这一结论的人是挪威数学家尼尔斯·亨里克·阿贝尔（Niels Henrik Abel），他于1824年发表了他的证明。

几乎在整个18世纪中，人们一直相信一般五次方程有代数解。找到这个解成了那个时代最难的问题。到了1700年，即费拉里找到四次方程的解160年之后，关于五次方程的问题还没有取得丝毫进展。适用于三次方程和四次方程的技术不能解决五次方程的问题。显然需要一种全新的想法。

而在此之前的17世纪和18世纪初这一问题被忽视了。实用的新文字符号体系的出现、微积分的发现、对复数的接纳以及理论科学的快速发展，使数学家们有了大量触手可得的成果。需要深入研究但又没有明显实际应用的难题在这样的环境下失去了吸引力。这样的问题是标准的（请允许我在这里用这个词）纯数学问题。对于给定的实际的五次方程，我们很容易找到它的数值解。

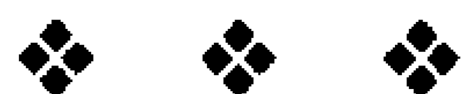


1732年，伟大的瑞士数学家欧拉首次接触一般五次方程的问题。当时他正居住在俄国的圣彼得堡，对此没有进行很深入的研究。但是30年后，他在柏林为菲特烈大帝工作时，在这个问题上有了更深一层的研究。在论文《论任意次方程解》中，欧拉指出任意n次方程的解的表达式也许是这样的形式：

$$A+B\sqrt[n]{\alpha}+C\left(\sqrt[n]{\alpha}\right)^2+D\left(\sqrt[n]{\alpha}\right)^3+\cdots$$

其中，α是某个$n-1$次“辅助”方程的解，而A、B、C等是原来方程系数的某个代数表达式。很好，但是又如何找到这个$n-1$阶的“辅助”方程的解呢？

这位18世纪最伟大的数学天才[1]（高斯是19世纪最伟大的数学天才）没有再深入研究下去。但是，欧拉的工作也不是毫无成果。1824年阿贝尔的五次方程不可解的证明就是从类似上面所给出的解的表示形式开始的。



有时候事情就是很出人意料，这种至关重要的见解居然不是18世纪最伟大的数学天才提出来的，而是一位无名小辈提出来的。

范德蒙其实是一个土生土长的法国人，从他的姓名上看不太出来。1770年11月他正好35岁，在巴黎法国科学院[2]宣读了一篇论文，随后他又在该科学院宣读了三篇文章（1771年他入选该科学院）。这四篇论文就是他的全部数学成果。他的主要兴趣好像是音乐，《科学传记辞典》中记载："据说，当时，音乐家认为范德蒙是一名数学家，而数学家又认为他是一名音乐家。"

范德蒙因以他的名字命名的行列式而广为人知（我将在后面介绍行列式），然而行列式实际上并没有在他的论文中出现，把这功劳归于他似乎是一个误解。总之，范德蒙是一个古怪又有点儿神秘色彩的人物，就像弗拉基米尔·纳博科夫①作品中的人物一样。后半生他成了一名雅各宾党人，狂热的法国革命支持者，1796年因病去世。

范德蒙的重要见解是用一个通用形式表示方程的每个解。例如，考虑二次方程$x^2+px+q=0$。假设它的解是α和β。下面的两个事实显然成立：

$$\alpha=\frac{1}{2}\left[(\alpha+\beta)+(\alpha-\beta)\right]$$

$$\beta=\frac{1}{2}\left[(\alpha+\beta)-(\alpha-\beta)\right]$$

稍微改变一下方式。考虑到平方根符号预示两个可能值，一正一负，那么下面表达式的两种可能值就是这个方程的解：

$$\frac{1}{2}\left[(\alpha+\beta)+\sqrt{(\alpha-\beta)^2}\right]$$

① 俄裔美国作家。他的作品致力于用语言制造扑朔迷离的时空迷宫，制造个人有别于"早已界定"的生活与现实，显示出了一种华美玄奥的风格。其代表作有《洛丽塔》和《微暗的火》等。——编者注

上面这个表达式的关键是什么呢？$(\alpha-\beta)^2$等于$(\alpha^2-2\alpha\beta+\beta^2)$，而这个表达式可以变成［$(\alpha+\beta)^2-4\alpha\beta$］，这是一个关于$\alpha$和$\beta$的对称多项式。记住，用解表示的对称多项式总是可以转变成用系数表示的多项式，上面的情况就是(p^2-4q)。

由此就可得到我们熟悉的二次方程的一般解。当然，这不是问题的关键点。关键点是这种方法好像可以一般化为对任意次方程都适用的方法，而前面解二次方程、三次方程和四次方程的方法都是特殊的，不能一般化。

让我们把上面的方法推广到三次方程$x^3+px^2+q=0$上。如通常一样用ω和ω^2表示两个复单位立方根，回想一下介绍单位根时所讲的内容，它们满足三次方程$1+\omega+\omega^2=0$，由此得出一般解：

$$\frac{1}{3}\left[(\alpha+\beta+\gamma)+\sqrt[3]{\left(\alpha+\omega\beta+\omega^2\gamma\right)^3}+\sqrt[3]{\left(\alpha+\omega^2\beta+\omega\gamma\right)^3}\right]$$

因为这是一个不完全三次方程，$(\alpha+\beta+\gamma)$等于零（参见第5章），所以只需考虑两个立方根项。但还有一个问题：与我处理二次方程时平方根符号下的表达式关于α和β对称这一情况不同，上式中立方根符号下面的表达式不是关于α、β和γ对称的。

再深入研究一下。我设$U=(\alpha+\omega\beta+\omega^2\gamma)^3$，$V=(\alpha+\omega^2\beta+\omega\gamma)^3$。那么在$\alpha$、$\beta$和$\gamma$的所有6种可能置换下，$U$和$V$将如何变化呢？用一种更容易看清楚的方式表示这些置换：

置换	U	V
$\alpha\to\alpha,\ \beta\to\beta,\ \gamma\to\gamma$：	$(\alpha+\omega\beta+\omega^2\gamma)^3$	$(\alpha+\omega^2\beta+\omega\gamma)^3$
$\alpha\to\alpha,\ \beta\to\gamma,\ \gamma\to\beta$：	$(\alpha+\omega^2\beta+\omega\gamma)^3$	$(\alpha+\omega\beta+\omega^2\gamma)^3$
$\alpha\to\gamma,\ \beta\to\beta,\ \gamma\to\alpha$：	$(\omega^2\alpha+\omega\beta+\gamma)^3$	$(\omega\alpha+\omega^2\beta+\gamma)^3$
$\alpha\to\beta,\ \beta\to\alpha,\ \gamma\to\gamma$：	$(\omega\alpha+\beta+\omega^2\gamma)^3$	$(\omega^2\alpha+\beta+\omega\gamma)^3$
$\alpha\to\beta,\ \beta\to\gamma,\ \gamma\to\alpha$：	$(\omega^2\alpha+\beta+\omega\gamma)^3$	$(\omega\alpha+\beta+\omega^2\gamma)^3$
$\alpha\to\gamma,\ \beta\to\alpha,\ \gamma\to\beta$：	$(\omega\alpha+\omega^2\beta+\gamma)^3$	$(\omega^2\alpha+\omega\beta+\gamma)^3$

（第一个置换不发生变化，叫作“恒等置换”。）

上面的置换好像没有给出更多的信息。然而，记住，ω是单位的立方根。我们可以利用这个事实把α项前面的ω和ω^2“分离出来”。例如，取第五行的第一项：

$$\left(\omega^2\alpha+\beta+\omega\gamma\right)^3=\left[\omega^2\left(\alpha+\omega\beta+\omega^2\gamma\right)\right]^3=\omega^6\left(\alpha+\omega\beta+\omega^2\gamma\right)^3$$

上面的括号里的结果正好是（第一行中的）U。事实上，其他每一个置换都可以转换成U或者V，正好有一半转换成U，一半转换成V。解的任意一种可能置换要么令U和V保持不变，要么令U与V互换。

为了让读者理解得更透彻，我再用略微不同的方式重述一次。我观察了所有可能置换，发现一半的置换保持U和V不变，而另一半置换把U变成V且把V变成U。

这里最关键的概念是对称。用解α、β和γ表示的多项式都是对称的，就像韦达和牛顿曾经得到的研究成果那样：用所有6种可能的方式置换解，多项式的值不变。它只有一个值。或者，一个多项式也可能是完全非对称的：用所有6种可能的方式置换解，多项式取6个不同的值。或如我的例子所示，多项式可能是部分对称的：用所有6种可能的方式置换解，这个多项式的取值数目大于1但小于6。

现在，对解进行操作：从所有这些事实可知，解α、β和γ的任意可能置换都将保持$(U+V)$和UV（或者任意其他U和V的对称多项式）不变。所以，$(U+V)$和UV本身一定是解α、β和γ的对称多项式，它们可以用这个三次方程的系数p和q表示。事实上，如果仔细观察这个代数式（并且记住在一个不完全三次方程中，x^2的系数为$\alpha+\beta+\gamma=0$），会发现：

$$U+V=-27q\,,\quad UV=-27p^3$$

所以，如果求解下面这个关于未知量t的二次方程：

$$t^2+27qt-27p^3=0$$

就会得到U和V。这个三次方程就有解了。（把这个解法与第3章后的“数

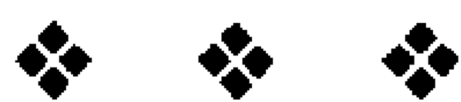
学知识：三次方程和四次方程”中的解法相比较。）

这个方法还有第二个问题。如以前一样，一般解的那个表达式中的根符号可以理解成容纳了这个根的所有可能值。例如，对于我们讨论的三次方程的情况来说，就是这个立方根的所有3种可能值：一个数、ω乘以这个数、ω^2乘以这个数。因为在我们的方括号里有两个立方根，所以这个表达式总共表示9个值：三次方程的3个解和其他6个不相关的数值。我如何求出这9个具体的值呢？

范德蒙实际上没有完全解决这个问题。但是他给出了一个非常重要的见解。对于三次方程来说：

> *把一般解写成所有解的对称多项式或者部分对称多项式形式。*
>
> *问：在α、β和γ的所有6种可能置换下，这个多项式有多少个不同的值？*
>
> *答案是两个，我们把它们称为U和V，于是就得到了一个二次方程。*

以上就是第一个尝试：通过观察解的置换及观察保持某个表达式不变的那些置换的子集来求解方程。在我的例子中，$(\alpha+\omega\beta+\omega^2\gamma)$的立方就是这个保持不变的表达式。以上就是攻克一般五次方程问题的中心思想。



可怜的范德蒙，若他的成就同另一位更伟大的天才相比就立即黯然失色了。爱德华教授如是说：“范德蒙是一个有非法国式名字的法国人，而拉格朗日是一个有法国式名字的非法国人。”[3]

1736年拉格朗日（Giuseppe Lodovice Lagrangia）[4]出生在距法国边境50公里的意大利西北部城市都灵。虽然他不是一个法国人，但是他有部分法国血统，而且似乎更喜欢用法语写作，他早年就开始用法语姓氏Lagrange写作。（他的法语带有很重的意大利口音。）1787年他成为法国

科学院会员，并一直生活在巴黎，于1813年去世。他经历了法国大革命，并为重量和度量的公制体系的制定做了大量重要工作，因此他以法文名字约瑟夫-路易·拉格朗日（Joseph-Louis Lagrange）闻名于世不无道理。现在巴黎有一个小公园就是以他的名字命名的，那里有这个城市最古老的一棵树。

拉格朗日早期的工作和生活是在都灵度过的，他16岁就在那里成为了数学教授。到了1770年范德蒙向法国科学院提交他的论文时，拉格朗日已搬到了柏林，来到菲特烈大帝的宫廷。事实上他是欧拉的继承人，在欧拉离开后的1766年进入宫廷。菲特烈大帝显然发现拉格朗日对当时的政治和哲学非常了解，而且头脑灵活、善解人意，比老实巴交的欧拉圆滑多了，所以非常高兴他能接替欧拉在自己身边工作。

拉格朗日对代数的伟大贡献发生在1771年，是在范德蒙向巴黎科学院提交论文的几个月之后。这一成果是由菲特烈大帝掌管的位于柏林的弗里德里克科学院发表的，其论文题目是"关于方程代数解的思考"。正是这篇由业已成名的数学家所写的文章把通过方程解的置换来求解方程的思想呈现给广大数学家。

这的确非常不公平，是范德蒙首先得出了这一结论的，现代课本已经正式认可了这一点。然而，他的论文没有得到应有的关注，据巴什马科娃和斯米尔诺娃说："对代数发展没有产生影响。"直到1774年这篇论文才发表，当时拉格朗日的论文已经广为人知。没有证据表明拉格朗日知道范德蒙的工作。总之，拉格朗日是一个值得信任的人，如果他知道这件事就应该会承认。这是两个伟大的头脑，或者更精确地说是一个伟大的数学大脑与一个出色的数学大脑，有相同想法的案例。

拉格朗日与范德蒙采用了相同的思路，但是前者数学功底更深厚，对问题的分析更透彻。我还是以一般的不完全三次方程$x^3+px+q=0$为例加以说明。

拉格朗日一开始与范德蒙使用相同的表达式$\alpha+\omega\beta+\omega^2\gamma$（我们称其为

拉格朗日预解式），但他注明，当置换解α、β和γ时，上面表达式取6个不同的值，尽管它的立方只取两个值（前文已说明）。这些值是：

$$t_1=\alpha+\omega\beta+\omega^2\gamma \qquad t_2=\alpha+\omega^2\beta+\omega\gamma \qquad t_3=\omega^2\alpha+\omega\beta+\gamma$$

$$t_4=\omega\alpha+\beta+\omega^2\gamma \qquad t_5=\omega^2\alpha+\beta+\omega\gamma \qquad t_6=\omega\alpha+\omega^2\beta+\gamma$$

同前面一样，我可以利用ω是1的立方根把它从α项提取出来。于是有，$t_3=\omega^2t_2$、$t_4=\omega t_2$、$t_5=\omega^2t_1$及$t_6=\omega t_1$。

现在，我们做一个六次多项式，使得这些t是它的解。这个多项式是：

$$(X-t_1)(X-t_2)(X-\omega^2t_2)(X-\omega t_2)(X-\omega^2t_1)(X-\omega t_1)$$

拉格朗日称其为*预解方程*。很容易把它简化成：

$$(X^3-t_1^3)(X^3-t_2^3)$$

这个多项式是我们前面得到的二次方程，它的解是$U=t_1^3$和$V=t_2^3$。

拉格朗日将同样的求解方法用在了一般四次方程上，这一次他得到了一个24次预解方程。与三次方程的预解方程“降次”成一个二次方程一样，这个四次方程的24次预解方程降次成一个6次方程。这看起来不太好解，但是关于X的6次方程实际上是一个关于X^2的三次方程，所以还是可解的。

5个对象的可能置换的数目是1×2×3×4×5=120。因此五次方程的拉格朗日预解方程是120次方程。用一定的技巧可以把这个预解方程降次成一个24次方程。就是在这里，拉格朗日被困住了。但是，同范德蒙一样，他也抓住了关键点：为了弄明白方程的可解性，必须观察它们的解的置换，以及在这些置换的作用下特定的关键表达式，即预解方程发生了什么样的变化。

他还证明了一个重要的定理，这就是今天我们告诉学习代数的学生的拉格朗日定理。我用拉格朗日自己对此的理解方式陈述这个定理。这个定理的现代形式与此略微不同，但更通用。

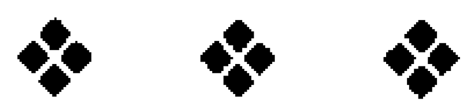
假设有一个n变量多项式[5]，那么就有1×2×3×…×n种置换这些变量的方法。你也许知道，这个数被称为“n的阶乘”，写成n!。所以，2！=2，3！=6，4！=24，5！=120，依此类推。（习惯上1!取1。但出于更全

面的考虑，规定0!也取1。）假设同前文处理α、β和γ一样，用$n!$种方法置换这n个变量，那么这个多项式有多少个不同的值呢？在前文中，这个答案是2，分别为U和V。但是一般来说这个答案能够是多少呢？如果某个多项式取A个不同值，我们能确定A有什么特征吗？

拉格朗日定理说，A总是一个可以整除$n!$的数。因此，使用α、β和γ构建任意的多项式，按所有6种可能形态更换3个未知量，并计算多项式能取多少个不同的值。如果多项式是对称的，那么答案是1。如果多项式是上面所给的例子，那么答案是2。如果这个多项式是$\alpha+\beta-\gamma$，那么答案是3。如果这个多项式是$\alpha+2\beta+3\gamma$，那么答案是6。但是它不可能是4或者5。这就是拉格朗日证明的事情。当然，他是对任意数值n进行证明的，而不只是对于$n=3$的证明。

（注意，拉格朗日定理给出了一个A的性质：它可以整除$n!$，但是不能保证每一个整除$n!$的数一定是一个可能的A，这是某个多项式在$n!$个置换下的可能取值的数量。例如，假设n是5，那么$n!$是120。因为4可以整除此120，你也许希望存在某个五变量多项式，在经过120种可能的置换之后，这个多项式取4个值。其实不是这样的。这一事实是由下一章将介绍的数学家柯西发现的，是寻找一般五次多项式代数解问题的关键。）

拉格朗日定理是现代群论的基础之一，而群论在拉格朗日时代还不存在。



本章出现的第一个名字是保罗·拉菲尼，他几次试图证明一般五次方程没有代数解。拉菲尼是拉格朗日思想的追随者。对于一般三次方程，我们能够得到它的一个二次预解方程，并且知道如何求解。对于一个四次方程，我们可以得到一个三次预解方程，同样知道这个三次方程如何求解。拉格朗日已经证明为了求一般五次方程的代数解，我们需要得到一个三次或四次的预解方程。拉菲尼仔细观察了置换未知量时多项式的可能取值，最后证明对于五次方程，不可能得到一个三次或四次的预解方程。

为拉菲尼[6]写传记的一位作家说："我们应该对拉菲尼深表同情和敬意。"的确如此。他的第一个证明有缺陷，但是他继续研究，至少发表了3个证明。他把这些证明寄给他那个时代的资深数学家，其中就包括拉格朗日，但是这些证明不是被忽视就是借口说不了解而拒绝接受，特别是拉格朗日的回复尤其令他感到痛苦，因为他认为拉格朗日和他是意大利同胞，应该认真对待他的研究成果。接着，拉菲尼又把他的证明提交给各学术团体，包括法国协会（替代因大革命而被暂时废弃的科学院）和柏林皇家协会，但结果都一样。

一直到1822年他去世，可怜的拉菲尼一直没有放弃让大家认可他对于证明的努力，但都无果。只是在1821年，对他的工作确有一份实实在在的认可。那一年，伟大的法国科学家奥古斯丁-路易·柯西给他寄来一封信。一直到去世前几个月，这封信都被拉菲尼珍藏着。柯西在这封信中赞扬了拉菲尼的工作，并声明以柯西自己的观点看，拉菲尼已经证明了一般五次方程没有代数解。事实上，柯西于1815年发表了一篇论文，显然就是以拉菲尼的证明为基础的。

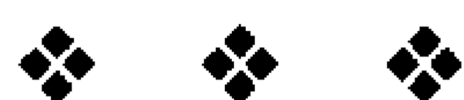
既然柯西这个名字再次出现了，我就介绍几句吧。在数学历史中，这是个伟大的名字。"以柯西命名的概念和定理比任何其他数学家都多。（大约有16个概念和定理是以柯西命名的。）"这段话摘自《科学传记辞典》中弗罗伊登塔尔为柯西所做的词条，这一词条长达17页，与高斯词条一样长。

然而，柯西的研究风格与高斯完全不同。高斯发表文章非常谨慎，只发表那些他已经弄清楚且非常完美的结果。（这就是为什么他的出版物非常难以理解的原因。）所以这150年来一直流传着这样一个关于数学家的笑话：当某人得出一个灿烂耀眼并且明显是原创的结果时，他还是要马上核查一下这个结果是否在高斯的某个未发表的论文中出现过。柯西就不一样，他头脑中闪现的每一个智慧的火花都会发表，而且是几天之内就发表。为此他创办了一个私人杂志。

对于柯西的个性有很多评论。不同的传记作家对他的描述也不同，

有人把他描绘成一个虔诚、正直和慈善的人，也有人把他描绘成一个冷血、善于专权的阴谋家，还有人把他描绘成一个白痴的专家，生活能力差、混乱不堪。柯西是一位虔诚的天主教徒，一位坚定的保皇者[7]，当欧洲的知识分子阶层以极大的热情投入非宗教的改良政治运动中时，他却极力反对。

针对历史上诸多对他不利的评论，现代人更倾向于对柯西给出正面的评价（参见第8章）。著名数学史家E・T・贝尔对柯西的评论也比较公平："他的性情很温和，除了数学和宗教之外，他对任何事都表现得很温和。"弗罗伊登塔尔倾向于认为他是一个白痴的专家的观点："说他有唐吉诃德式的举止着实有些夸张，人们更倾向于认为他是情绪不定……柯西看起来就像个孩子似地天真。"这个人的个性如何另当别论，但是他是一位非常伟大的数学家，这一点不容置疑。



在这样的情况下，我们也许会同情可怜的拉菲尼，而轻视阿贝尔，对他获取证明一般五次方程代数不可解性的荣誉有些不屑。

事实上，没有人这样想。首先，尽管柯西认为拉菲尼证明了这一问题，但是他那个时代的数学家都认为他的证明是不够严谨的。（现代人对拉菲尼更友善些，有时候一般五次方程的代数不可解定理也被称为阿贝尔-拉菲尼定理。）其次，这些证明的书写风格令人难以理解，这正是拉格朗日否认他的原因。再次，尽管阿贝尔是一个乐观、善于交际的人，但他的人生境况比拉菲尼更加可怜。

阿贝尔是19世纪挪威代数学三巨头的第一人。在后面我将介绍另外两个人。他的出生地是欧洲北部边缘地带，靠近斯塔万格，在挪威版图的"鼻子"上，经常受强风袭击。这个地方本身就很穷，时局不稳就使它更穷、更不幸。阿贝尔的家庭属于上流社会中的穷人，他的父亲和祖父都是当地的牧师。阿贝尔的父亲政治前途不顺，酗酒，"死于酒精中毒，留下九个孩子和一个寡妇。这位母亲也为寻求安慰而酗酒。丈夫丧礼后

牧师来访时，她与她的农夫情人竟然仍在床上鬼混。”[8]

阿贝尔死于他27岁生日的几周之前。在其短暂的人生中，过得最好的时候是捉襟见肘，不好的时候更是债台高筑。他家乡的境况也跟他的人生遭遇类似。到了阿贝尔十几岁的时候，挪威还是一个半独立国家，是挪威与瑞典联合王国的一部分，首都是奥斯陆，当时叫克里斯蒂安娜，有自己的议会，但却生活在相对富足、人口众多的瑞典的经济和军事阴影下。在这种情况下，挪威人能在1825年～1827年筹集足够多的基金送这位不知名的年轻数学家去欧洲真是值得赞扬。但这些基金的短缺和支出受到的管制引起了为阿贝尔写传记的作者的不满，甚至后来的挪威政府也为此感到内疚。

阿贝尔很早就接触到了数学，并非常幸运地受到霍尔姆博老师的指导，这位老师非常欣赏他的才能。尽管霍尔姆博自己不是有创造能力的数学家，但是他知道当时数学界的主流话题。得到霍尔姆博的鼓励和经济上的帮助，阿贝尔于1821年到1822年在克里斯蒂安娜大学求学。

从1820年开始，阿贝尔就已经开始研究一般五次方程的问题，并已经证明了不可解定理。1824年他自己出钱印刷发表了这一结果。为了节约开支，他把这一证明压缩到只有6页纸，因此牺牲了证明中的很多细节部分。尽管如此，他肯定这6页纸就能为他打开欧洲最伟大数学家之门。

当然，事情并没有像他预料的那样。在阿贝尔准备亲自拜访高斯之前，给高斯寄了一份自己的证明，但是伟大的高斯看都没看。我的意思不是说高斯是个不怎样的家伙。因为高斯已经很有名望，而且是一位著名的数学家，因此经常遭受一些声称自己已经证明了某某著名问题的行为古怪之人的骚扰。[9]在其声望鼎盛时期，高斯很不高兴被人愚弄，而且似乎对多项式方程代数解的问题也不太感兴趣。所以阿贝尔放弃了拜访高斯的计划。

似乎是老天有意给阿贝尔一份补偿，他在柏林有了意外的好运。他遇见了克瑞勒，一位数学史中独一无二的人物。克瑞勒（1780—1855）

不是一名数学家，而是一名数学经纪人。他善于发现数学天才和优秀人才，而且当他发现人才时，会倾其所有来帮助他们。克瑞勒出身卑微，全靠自力更生，基本是自学成才。他是普鲁士政府的土木工程师，而且干到了这一职业的最高级别。1838年他负责从柏林到波茨坦的德国第一条铁路的施工。克瑞勒是一位友善、慷慨而且精力充沛的人，他充当了伟大数学天才的助产师，间接对19世纪的数学做出了巨大的贡献。

就在1825年阿贝尔来到柏林时，克瑞勒刚打算创办属于自己的数学期刊。他发现了这位年轻的挪威数学天才（显然他们用法语交流），把他介绍给柏林的每一个人，在刚创办的《纯数学和应用数学》期刊的第一期发表了他的不可解证明。他还发表了更多阿贝尔的论文。五次方程不可解证明只是阿贝尔广泛数学兴趣中小小的一部分。他主要研究的是分析学和函数论。

1827年春天，身无分文的阿贝尔回到了挪威，从此再没有离开挪威。1829年他因肺结核去世，那个时代肺结核是不治之症。阿贝尔去世两天后，克瑞勒写信告诉他柏林大学已聘他为教授，当然那时他不知道阿贝尔已经不在人世了。

阿贝尔的证明整合了他掌握的欧拉、拉格朗日、拉菲尼和柯西等人的思想，并利用其独创的方法和见解将这一切完整地结合在了一起。这一证明的一般形式是反证法的典范：从假设想要的结果不成立开始，然后证明这将导致逻辑矛盾。

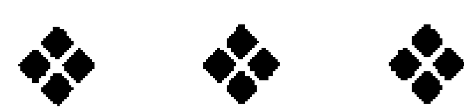
阿贝尔想要证明的是一般五次方程不存在代数解。因此他开始假设存在某个解。他的五次方程的一般形式是：

$$y^5-ay^4+by^3-cy^2+dy-e=0$$

他的证明过程如下。假设存在一个代数解，所有这样的解y都可以用a、b、c、d和e的表达式表示，而这些表达式只包含有限次加、减、乘、除和开方。当然，这些开方可以是“嵌套的”，就像一般三次方程的解中立方根之下的平方根那样。所以，我们用某个一般却（对求解）有效的

方式表示这些结果，其中的量可以有开方，其下又可以有开方，等等，即允许这样的嵌套。

阿贝尔给出一个一般解的表达式，这一表达式与我在前文中给出的欧拉的表达式非常相似。借用拉格朗日的表达式（和范德蒙的，尽管阿贝尔不知道他的表达式），他主张这个一般解一定能表示成为所有解和5个单位根的多项式。然后，阿贝尔利用了前文中介绍的柯西的结果：对于一个有5个未知量的多项式，当置换这些未知量时，它取2个不同值，或者是5个不同值，但不能是3个或者4个不同值。把这一结果运用到他的解的一般表达式，就得到了他需要的逻辑矛盾。[10]



阿贝尔的证明（或更严格地说是阿贝尔-拉菲尼证明），即一般五次方程没有代数解的证明，开辟了代数史上第一个伟大的新纪元。

在结束这一新纪元的介绍时，我要对这一事件发表一些事后的看法。在1826年，没有人是这样理解的。事实上，阿贝尔的证明是经过了一段时间之后才被人广泛认知。这一证明发表9年后的1835年，在都柏林举行的英联盟会议上，数学家杰拉德宣读了一篇论文，在这篇论文中他说自己已经找到一般五次方程的一个代数解！20年后，他仍然坚持这个观点。

阿贝尔的证明并没有终结一个未知量的多项式方程的一般理论。即使一般五次方程没有代数解，但特殊的五次方程有解。例如，在前面介绍单位根时，我已给出方程$x^5-1=0$的一个解，它毫无疑问是五次方程。所以就有了这样的问题：哪些五次方程可以代数求解，而且这些解可以用加、减、乘、除、开方符号以及五次方程系数组成的多项式表示？对这个问题给出完整答案的是另一位法国数学天才伽罗华，我将在第11章介绍他的故事。

然而，在19世纪的前几十年间，人们对代数的认识发生了重大但缓慢的转变。在阿贝尔印刷发表他的6页证明很久之前，这样的转变就已经开始了。我把这种新的思维方式定义是“新数学对象的发现”。而在此之

前，整个18世纪到19世纪初期，代数还只被认为是牛顿著作的标题所称呼的东西：泛算术。它是利用符号的算术，是对数值进行运算。

但这些年间，欧洲数学家一直在吸收同化17世纪的大师们留给他们的深入其心的完美符号体系。渐渐地，附加了这一体系的数值世界的束缚越来越少，数值可以随意组合，“过上了自己的生活”。就像两个数值加起来得到一个新数值那样，是否存在其他可以组合的东西，使得两个这样的东西结合到一起形成同一类型的另一个对象呢？显然已经存在这样的东西。高斯在他1801年的经典著作《算术研究》中就讨论了二次形式，即两个未知量的多项式，如下所示：

$$AX^2+2BXY+CY^2$$

研究之后，高斯有了合成这样的形式的想法，即把两个二次形式组合起来得到一个新的二次形式，这要比表达式的简单加法和乘法精妙得多。如高斯写到的，这是“一个还没有人考虑的课题”。

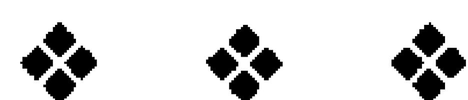
接下来就有了1815年柯西关于置换多项式的未知量时它的取值数量的论文。这就是阿贝尔在其证明中引用的论文。在这一论文中，柯西介绍了复合置换的思想。

下面给出一个简单的例子。假设我有3个未知量α、β和γ，而且假设给出切换β和γ的置换X、切换α和β的置换Y。现在假设我首先做置换X再做置换Y。会发生什么事情呢？置换X把(α,β,γ)转变成(α,γ,β)，然后置换Y把(α,γ,β)转变成(β,γ,α)。所以综合起来的效应是把(α,β,γ)转变成(β,γ,α)，这就是另外一个置换！我们可以称其为置换Z，合成置换X和置换Y得到一个置换Z。这就是柯西描述的运算方法，因此应该说柯西实际上发明了群论（只是他没有使用这个术语）。

这样一来，柯西就进入了一个奇妙的新世界。记住，举个例子，合成置换与数值相加的相似性是如何在一个重要方面被打破的。如果先做置换Y，然后做置换X，结果就是(γ,α,β)。所以，合成时的顺序是有讲究的。数值相加没有这样的情况，(7+5)等于(5+7)。数值的这一特殊性质在

技术上称为交换性。柯西的置换合成是非交换的。

在19世纪早期，所有这一切都是未确定的。经过六代人对韦达和笛卡儿的符号体系的研究之后，数学家开始认识到把两个数值通过加法或乘法相结合而得到一个新数值的合成只是这类运算的一种具体情况，它有更广泛的应用，可以运用到根本不是数值的对象上。他们经常使用的那些符号可以代表任意事物：数值、置换、数组、集合、旋转、转换、命题等。当这一点渗透进来时，现代代数就诞生了。



在接下来的几章中，我将不再严格按照年代顺序展开叙述。历史已经进入一个新代数思想多产的时期，也就是19世纪中间的五十年。在这一时期不仅发现了群，还发现了其他新的数学对象。“代数”已经不再是单数名词，而变成了一个复数名词。“域”、“环”、“向量空间”、“矩阵”等现代概念已经形成。乔治·布尔给代数符号体系赋予逻辑，而几何学家发现有比三维空间更大的空间需要他们去揭示，而这正是代数的功劳。

研究这一快速发展的史学家有两种选择：或者严格按照年代顺序，尝试着揭示新思想是如何产生的，同时揭示随着时间的推移新思想之间的相互作用；或者抓住这一时期的某一个思想线索，然后再循环往复抓住其他线索。我将采用后面的陈述方式，把代数飞速发展和代数思想的根本变化分成几条线路。首先介绍四维空间。

注　解

[1] 威廉·邓汉在《欧拉，我们所有人的大师》（1999年出版）一书中对这位伟人和他的数学研究做了公正的评价。

[2] 法国科学院于1666年由让·巴蒂斯特·柯尔培尔创建于巴黎，是17世纪后期欧洲科学伟大复苏的一部分，可与1660年创建的英国皇家学会相媲美。法国科学院院士曾经在卢浮宫集会。

[3] 见《伽罗华理论》一书第19页。

[4] 拉格朗日是一个伟人，按照查理·莫瑞的评分，它的指标积分是30（见

2003年出版的《人类成就》一书）。欧拉是这一领域的带头人，积分是100。牛顿89，欧几里得83，高斯81，柯西34。可怜的范德蒙的指标积分是1，这可能是根据“他的”行列式而给出的。

[5] 用“多项式”是为了简单起见，实际上是错误的，我应该说成是“有理函数”。介绍域论时我再解释其中的原因。用“多项式”暂时没有问题。

[6] 拉菲尼是医生，又是数学家。因此请允许我介绍一下另外一位同样有双重职业的18世纪代数学家。这个人就是爱德华·华林（1736—1798）。1760年华林取代艾萨克·牛顿成为剑桥大学卢卡斯数学教授。七年后，就在他还担任这一职位时，他获得了医学博士学位。他似乎没怎么行过医。他在1762年的著作《分析杂论》中讨论了方程解的对称函数与这个方程的系数之间的关系，对此我已经在牛顿的笔记中讨论过。（华林这本著作的第二版被误称为《代数沉思录》。）我真够啰嗦的，请原谅，我还要介绍一位人物：瑞典数学家艾尔德兰·布林格。他在1786年指出，任意一个五次方程都可以化简成一个没有未知量的二次幂、三次幂和四次幂的方程，换句话说，可以化简成一个形如$x^5+px+q=0$的方程。我喜欢称这个方程为“超简化五次方程”。

[7] 柯西似乎真相信中世纪君权神授理论。该理论经常被误认为是新教的教义，但事实上这要回溯到中世纪，它在17世纪的法国很流行。如果这一说法是真的，柯西一定是坚信这一理论的最后一名声名显赫的知识分子。

[8] 这一故事摘自彼得·佩希奇的小册子《阿贝尔的证明》（2003年出版）。但E·T·贝尔说阿贝尔有七个孩子。另外，贝尔在《数学男人》一书中关于阿贝尔的介绍就像20世纪30年代美国史一样精彩，绝对值得一读。这是春风得意的阿贝尔，也是落魄失意的阿贝尔，就看你怎么看待作家们的评论了。

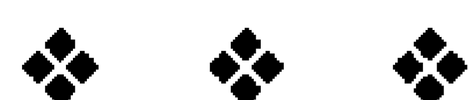
[9] 不是数学家也会受到骚扰。我的关于黎曼假设的书出版之后，我收到了大量的信件，都是那些声称自己已经解决了那个深奥谜题的人寄来的。我没时间仔细看他们的东西，又不愿意伤害他们，所以就准备了如下的套话常备回复：“我不是一位职业数学家，只是一个有数学背景的作家。我写了一本关于黎曼假设的书并不能说明我有资格评判这个领域的工作。我还曾经写过一本关于歌剧的书，但是我不会唱歌剧。我建议你与当地大学的数学系联系。”

[10] 关于阿贝尔的五次方程一般解的详细证明远远超出了本书的深度。我建议好奇的读者去看一下彼得·佩希奇的《阿贝尔的证明》。这本书尽可能使用初等的方法给出了这一证明，即用三段论递进解释：概要（比我的详细）、阿贝尔1824年的论文原文以及对此论文中没有写出的逻辑步骤的解释。范德瓦尔登的《代数的历史》也给出了一个两页半的概括，当然是在更高的层次上的。

数学知识

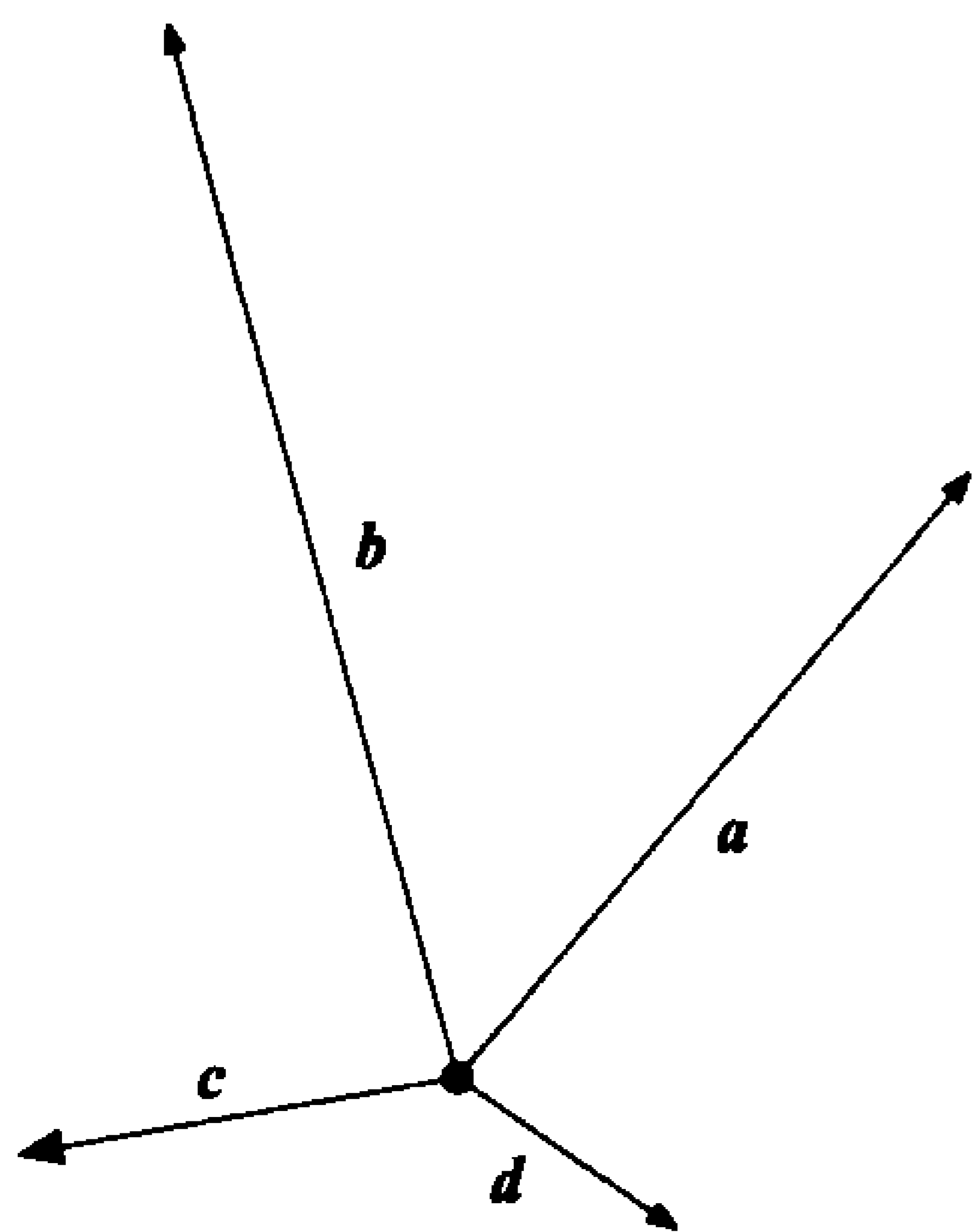
向量空间与代数

数学中向量这个概念的来龙去脉相当混乱，我在正文中将尽量把它理清一下。本部分介绍这一知识完全采用现代处理方法，使用大约1920年开始流行的概念和术语。



向量空间是一个数学对象的名称。这一对象有两种要素：向量和标量。标量可能是我们熟悉的某个数系，有加法、减法、乘法和除法，实数系 $\mathbb{R}$ 就是一个标量集。向量稍微有点复杂。

来看一个非常简单的向量空间实例，一个无限的平面。我选择这个平面上的一个特殊点，称其为*原点*。一个向量是从原点到另一个点的直线。图VS-1中给出了一些向量。可见，向量的两个特性是它的长度和方向。

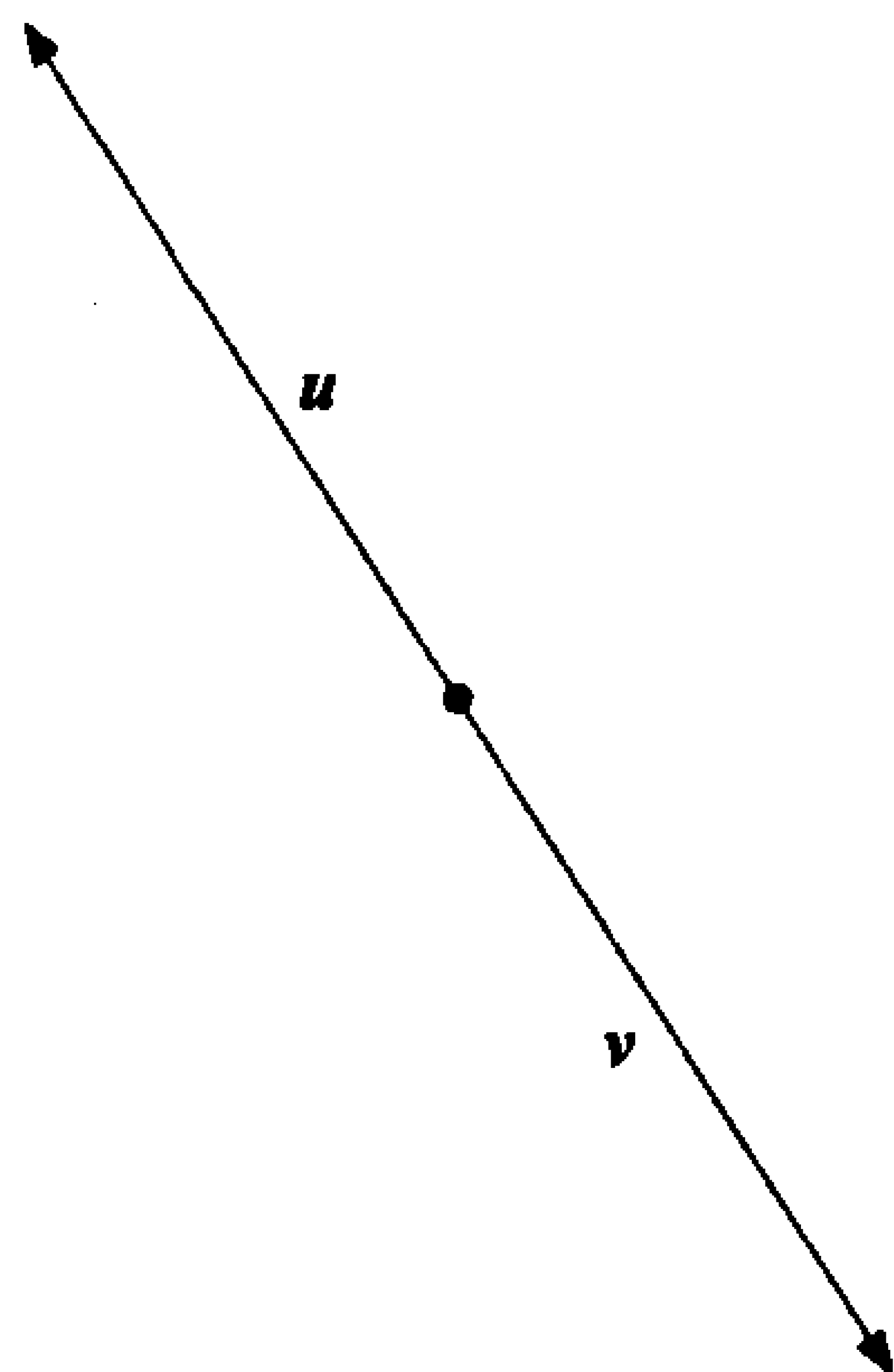


图VS-1　某些向量

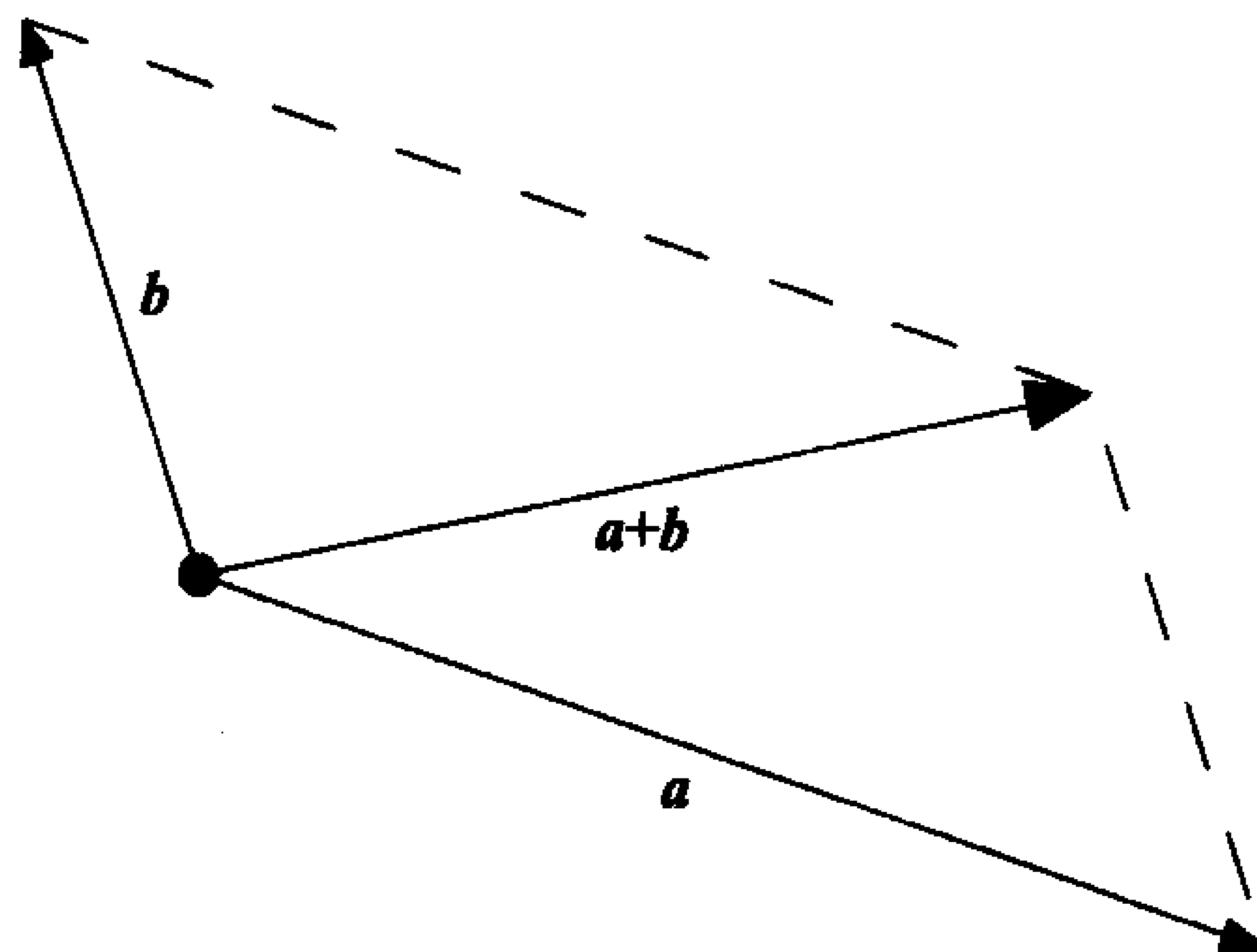
向量空间中的每一个向量都有一个逆向量。这个逆向量是与这个向量长度相同但指向相反方向的直线（见图VS-2）。原点本身可以看成是一个向量，称为*零向量*。

可以把任意两个向量加起来。为了把两个向量相加，把它们看成一个平行四边形的两条相邻的边。补画出这

个平行四边形的另外两条边。从原点到这个平行四边形远角的对角线就是这两个向量的和（参见图VS-3）。



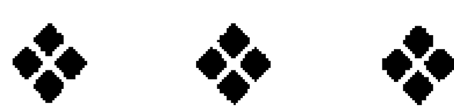
图VS-2 逆向量



图VS-3 向量相加

如果把一个向量与零向量相加，那么结果就是原来的向量。如果把一个向量与它的逆向量相加，结果是零向量。

任意一个向量可以与任意一个标量相乘。这个向量的长度发生了相应的改变（例如，如果这个标量是2，那么这个向量的长度加倍），但是它的方向不变。如果所乘的数是一个负数，那么这个向量被逆转了。



向量空间差不多就这些内容。当然，我在这里给出的平面向量空间只是一个例子，还有更多的向量空间，过一会再介绍，暂时还是用这个示例。

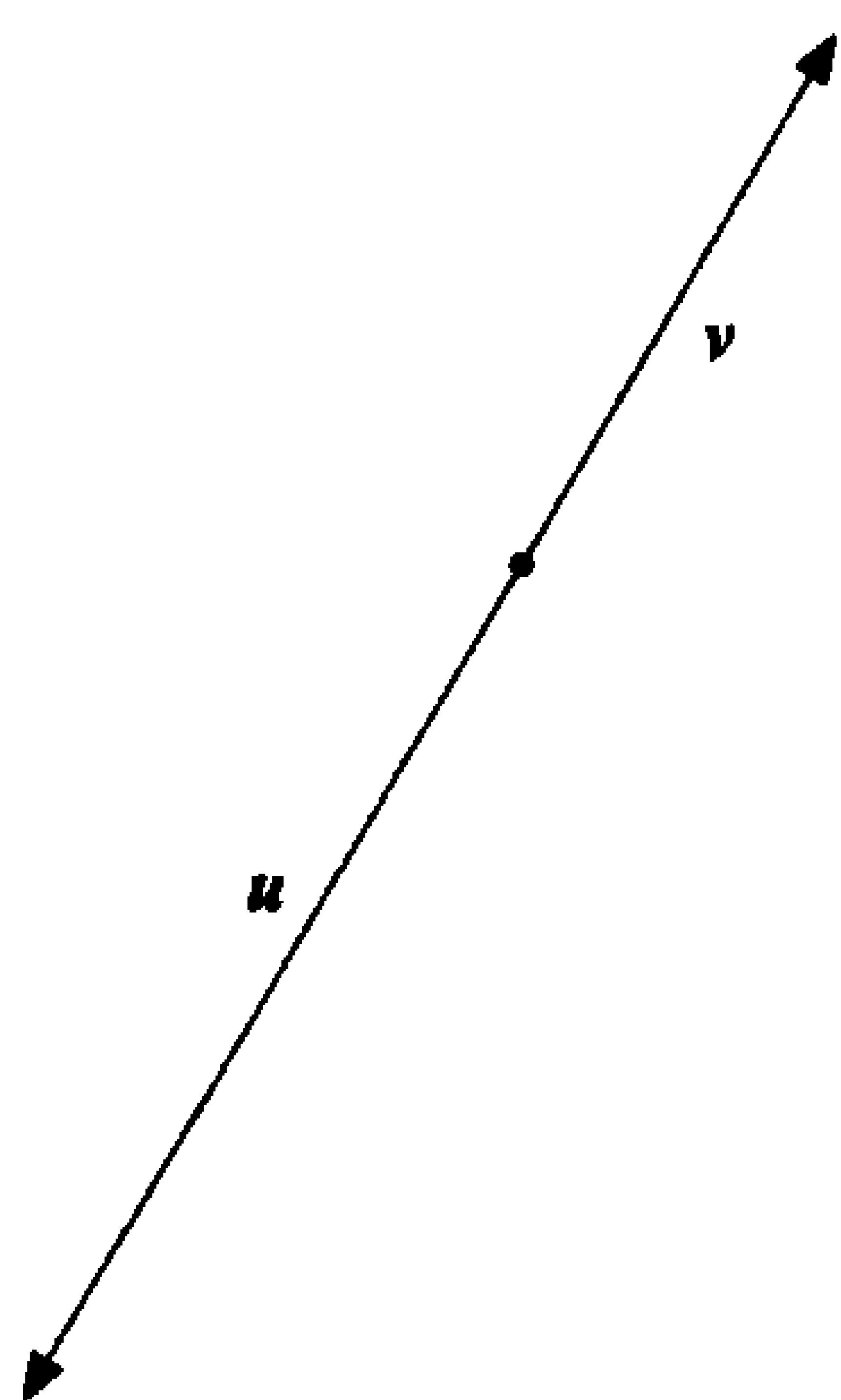
向量空间理论中的一个重要概念就是线性相关。在向量空间中任意取出一组向量，比如说$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w}$等。如果能够找到一组不都是零的标量p、q、r等，满足

$$p\boldsymbol{u}+q\boldsymbol{v}+r\boldsymbol{w}+\cdots=0 \text{（即零向量）}$$

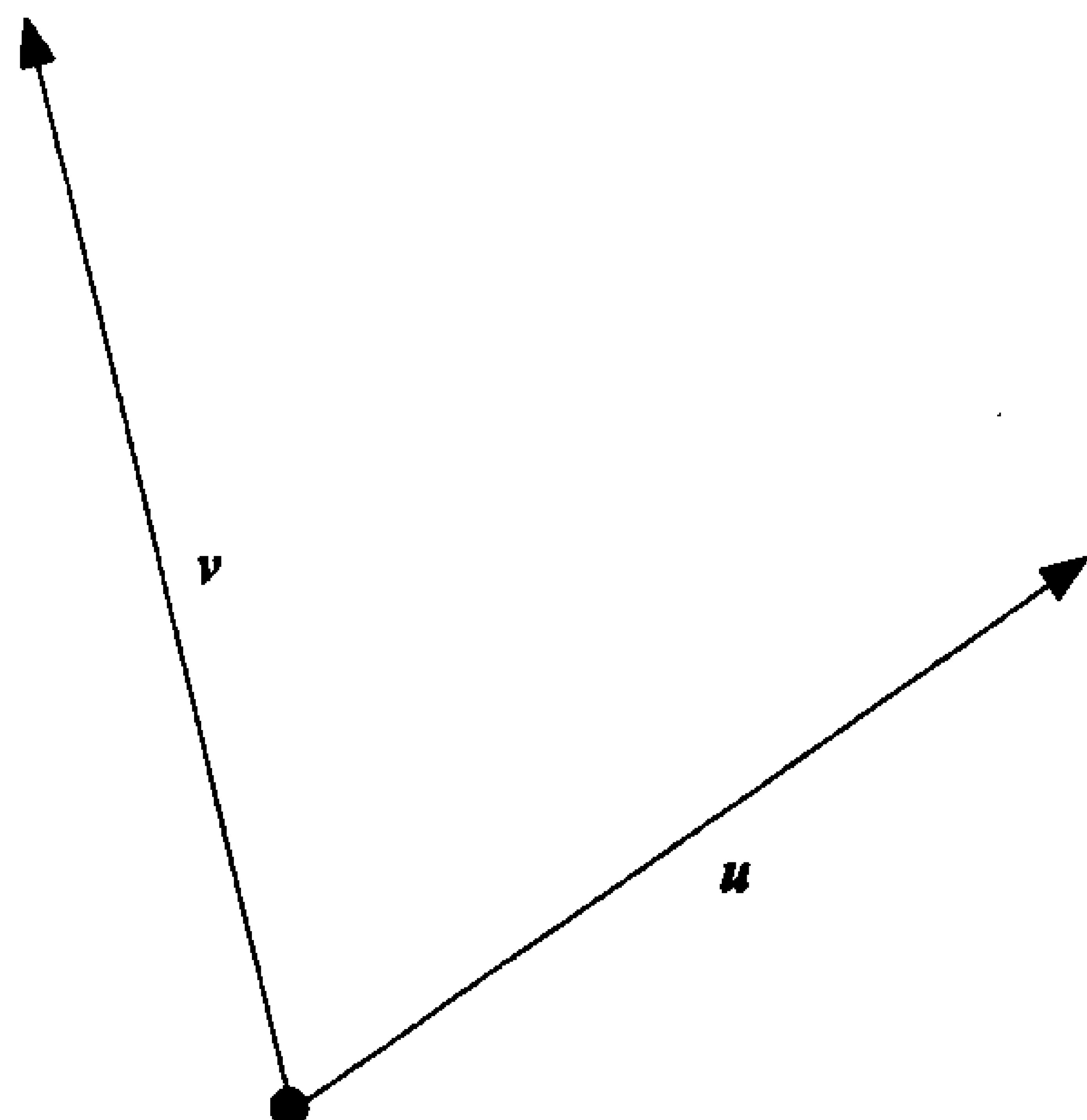
那么我们说向量$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w}$等是线性相关的。看一下图VS-4中的两个向量$\boldsymbol{u}$和$\boldsymbol{v}$。$\boldsymbol{v}$指向$\boldsymbol{u}$的相反方向，其长度是$\boldsymbol{u}$的2/3。于是有$2\boldsymbol{u}+3\boldsymbol{v}=0$。所以$\boldsymbol{u}$和$\boldsymbol{v}$

是线性相关的。还有另外一种表示这种线性相关性的方法：用其他向量表示其中的一个向量，$\boldsymbol{v}=(-2/3)\boldsymbol{u}$。

现在看图VS-5中的向量$\boldsymbol{u}$和$\boldsymbol{v}$。这一次这两个向量不是线性相关的。不可能找到不都为零的标量a和b，使得$a\boldsymbol{u}+b\boldsymbol{v}$等于零向量。[证明方法如下。因为$a$和$b$不能都是零，所以可以用其中的一个除以这个表达式，比如用b去除，于是把这个表达式变成这样的形式：$c\boldsymbol{u}+\boldsymbol{v}$。现在想象一下向量$\boldsymbol{u}$和$\boldsymbol{v}$相加的图形，它有图VS-3所示的平行四边形和对角线。保持$\boldsymbol{v}$不变，通过变化c的所有可能值来任意改变$\boldsymbol{u}$向量的大小，既可以把它变大或变小，也可以把它变成零或负的（当然对于c的负值，就是倒转它的方向），然后观察这个对角线所发生的变化。它永远不等于零。]



图VS-4　线性相关



图VS-5　线性无关

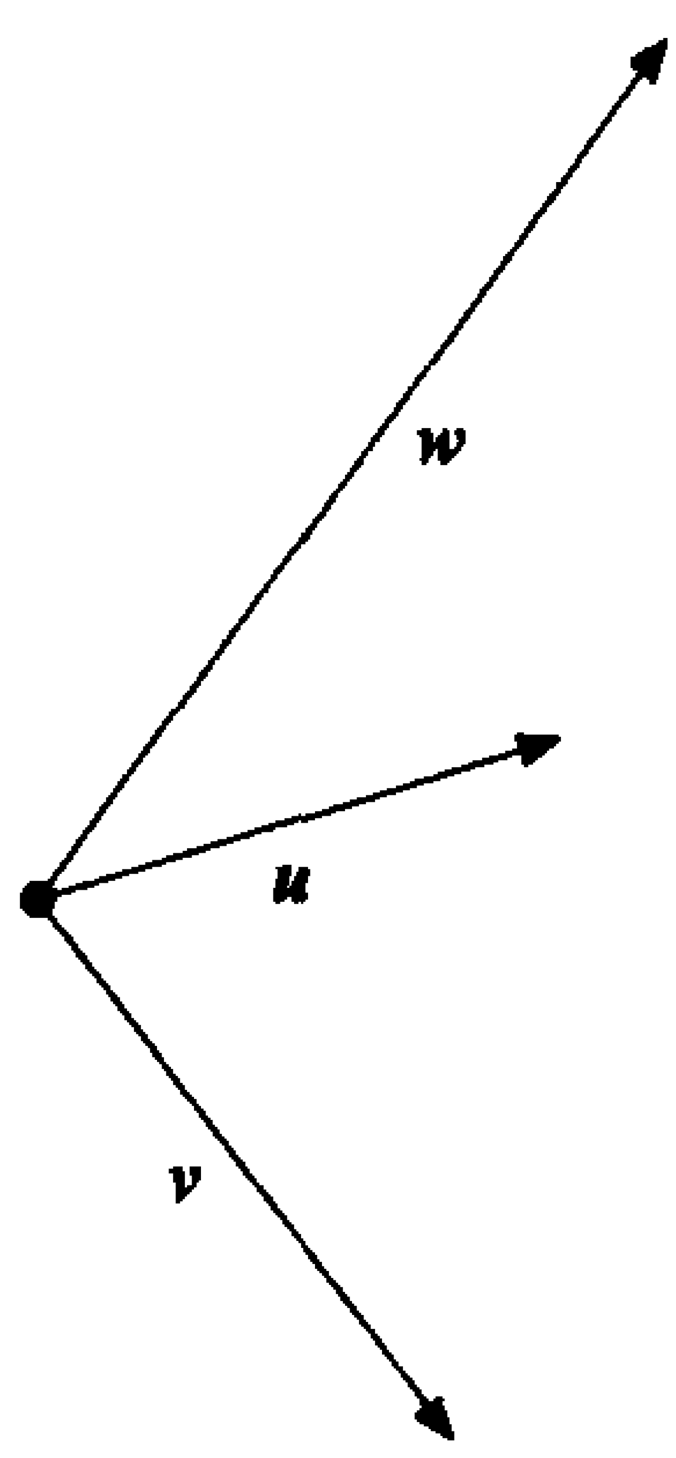
因此，图VS-5中的向量$\boldsymbol{u}$和$\boldsymbol{v}$不是线性相关的。它们是线性无关的，不能用其中的一个向量表示另一个向量。而在图VS-4中，可以用其中一个向量表示另一个向量：$\boldsymbol{v}=(-2/3)\boldsymbol{u}$。

有了向量集合的线性无关的概念，就可以定义向量空间的维度。它就是在这个空间中能够找到的线性无关向量的最大数目。在我们的向量空间示例中，可以找到很多两个线性无关的向量例子，如图VS-5中的两

个向量；但是找不到三个线性无关的向量。图VS-6中的向量$\boldsymbol{w}$可以用向量$\boldsymbol{u}$和$\boldsymbol{v}$表示，事实上我已经给出表示方法：$\boldsymbol{w}=2\boldsymbol{u}-\boldsymbol{v}$。换一种表示方法：$2\boldsymbol{u}-\boldsymbol{v}-\boldsymbol{w}=0$。向量$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w}$是线性相关的。

因为在示例向量空间中我能够找到的线性无关向量的最大数是2，所以这个空间的维度是2。我想你不会对这个结果感到很惊讶。

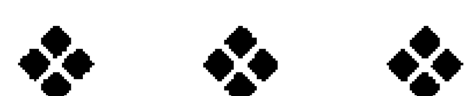
在一个像这样的二维向量空间中，如果两个向量是线性无关的，那么它们就不会在一条直线上。在一个三维空间中，如果三个向量是线性无关的，那么它们就不会在同一个平面内。反之，位于同一个平面，也就是相同的二维空间的三个向量一定是线性相关的，位于同一条直线的两个向量，相同的一维空间，它们一定是线性相关的。在相同的三维空间中的任意四个向量是线性相关的……依此类推。



图VS-6　$\boldsymbol{w}=2\boldsymbol{u}-\boldsymbol{v}$

给定两个线性无关的向量，例如图VS-5和图VS-6中的向量$\boldsymbol{u}$和$\boldsymbol{v}$，我的例子中的向量空间中的任意一个向量都可以用它们表示，就像前面表示$\boldsymbol{w}$那样。这说明这两个向量是这个空间的基。任意两个线性无关的向量都可以作为基，这个空间的任意一个向量都可以用它们表示。

维度和基是向量理论的基本术语。



向量空间的概念是纯代数的，不需要有任何几何表示。

考虑某未知量x的次数不大于5且系数取自$\mathbb{R}$的所有多项式。下面是一些这样的多项式的例子：

$$2x^5-8x^4-x^3+11x^2-9x+15$$
$$44x^5+19x^3+4x+1$$
$$x^2-2x+1$$

这些多项式是我的向量。我的标量是$\mathbb{R}$。这是一个向量空间。看，我可以把两个向量加起来得到另一个向量：

$$(2x^5-8x^4-x^3+11x^2-9x+15)+(44x^5+19x^3+4x+1)$$
$$=46x^5-8x^4+18x^3+11x^2-5x+16$$

每一个向量都有逆向量。（就是改变符号。）实数零充当零向量。当然，我还能把一个向量与一个标量相乘：

$$7\times(44x^5+19x^3+4x+1)=308x^5+133x^3+28x+7$$

这些公式都成立。很显然，x^5、x^4、x^3、x^2、x和1是这个空间的基。这些向量是线性无关的，因为只有当a、b、c、f、g、h全都是零时，对于未知量x的所有值，才有以下等式（等号右边为零的多项式）成立：

$$ax^5+bx^4+cx^3+fx^2+gx+h=0$$

而且因为这个空间中的其他任意向量都可以用这6个向量表示，没有7个或更多向量是线性无关的，所以这个空间的维度是6。

你也许抱怨说我没有认真处理这些多项式，例如，没有做因式分解，x的幂只是定位而已。在这里，我的操作对象就是其系数的六元组。难道不能忽略x并把例子中的多项式写成下面的六元组吗？

$$(2,-8,-1,11,-9,15)$$
$$(44,0,19,0,4,1)$$
$$(0,0,0,1,-2,1)$$

如果我接下来用下面很直观的方法定义六元组的加法：

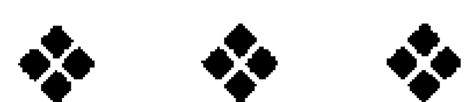
$$(a,b,c,d,e,f)+(p,q,r,s,t,u)=(a+p,b+q,c+r,d+s,e+t,f+u)$$

以及与标量的乘法：

$$n\times(a,b,c,d,e,f)=(na,nb,nc,nd,ne,nf)$$

那么在某种意义上，这不是与多项式向量空间相同的向量空间吗？

是的，它是。数学对象“向量空间”只是一种工具和一种抽象。我们可以用不同的方式表示它，例如用几何方式或多项式形式，这样就对使用它执行的特定任务会有更深刻的了解。但是，事实上，$\mathbb{R}$上每一个六维向量空间本质上与上面我给出的带有向量加法和标量乘法的实数六元组向量空间相同（数学上称为“同构”）。



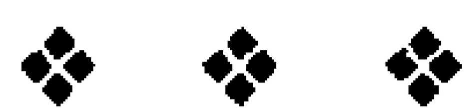
尽管向量空间本身并不那么令人兴奋，但是当我们研究它们之间的关系或者在基础模型上添加一些新特性来增强它们时，它们就会带来一系列更强大更迷人的结果。

内容最多的话题就是线性变换。所谓线性变换就是从一个向量空间到另一个向量空间的可能映射，根据某个定义规则，把第二个向量空间的一个向量指定给第一个向量空间中的一个向量。术语“线性”的意思就是这些映射有很好的行为，例如，如果向量$\boldsymbol{u}$映射到向量$\boldsymbol{f}$，而向量$\boldsymbol{v}$映射到向量$\boldsymbol{g}$，那么就能保证向量$(\boldsymbol{u}+\boldsymbol{v})$映射到向量$(\boldsymbol{f}+\boldsymbol{g})$。

如果你想把一个较高维的向量空间映射到一个较低维的向量空间，就能明白我所说的意思，这叫作“投影”。反之，较低维的空间也可以映射到较高维的空间，这叫作“嵌入”。我们还能把一个向量空间映射到它本身或者映射到它本身的一个较低维的子空间。

像$\mathbb{Q}$或者$\mathbb{R}$这样的数域就是一个其自身上的一维向量空间（如果没

有理解，请停下来思考一下），甚至可以把一个向量空间映射到它自己的标量域。这种映射称为**线性函数**。使用我们的多项式向量空间，就可以得到这种类型的映射例子，具体做法是在每一个多项式中用某个固定的数值替换x，例如x等于3。这样就把每一个多项式变成一个数值，所以是线性的。令人惊讶的是（我似乎总感到惊讶），一个空间上的所有线性函数集合自身形成一个向量空间，其函数就是向量。这个向量空间是原来向量空间的对偶向量空间，与原来空间有相同的维度。为什么在这里停止呢？为什么不把一对向量映射到基础域上，使得任意对$(\boldsymbol{u},\boldsymbol{v})$映射到一个标量呢？这样就得到了机械物理和量子物理专业的学生们所熟悉的概念：**内积**（或标量积）。我们甚至能够把向量的三元组、四元组、n元组映射到标量，这样就进入了张量、格拉斯曼代数以及行列式的理论范畴。虽然谦逊的向量空间本身是一个简单的东西，但它为我们打开了数学奇境的宝库。



对于向量之间的关系就讲这么多。新特性又是什么呢？如果略微增强一个向量空间，向这个基础模型加入一些新特性，那么能够得出什么理论呢？

用得最多的这类附加性质是**向量相乘的某种方法**。回想一下在向量空间的基本定义之下，标量形成一个域，它有加法、减法、乘法和除法四则运算。然而，对于向量只能进行加法和减法。标量可以与向量相乘，但向量之间不能做乘法。增强向量空间的一个很显然的方法就是加入某种合适且有效的向量乘法：两个向量相乘得到另一个向量。

有这种额外特性的向量空间被称为一个代数（an algebra），即在这个向量空间中，两个向量不仅可以相加，而且还可以相乘得到另一个向量。

我承认这不是一个非常令人高兴的用途。单词“代数”已经有了完美的含义，这就是本书的话题。为什么非要把一个不定冠词放在代数一

词的前面并用它给这一新数学对象命名，从而混淆这一议题呢？然而，抱怨是没用的。现在，这一用法已经被一致接受了。如果你听到某个数学对象被称为是“一个代数”，那么几乎可以肯定这是一个向量空间，向量相乘在其中成立。

“一个代数”的最简单且非常常见的例子就是复数空间$\mathbb{C}$。回想一下，我们可以这样形象地描述复数：在一个无限平面上铺开它们，实数部分就是东西坐标，虚数部分就是南北坐标（参见图NP-4）。这就意味着可以把复数空间$\mathbb{C}$看成一个二维向量空间。一个复数就是一个向量，标量域就是$\mathbb{R}$；零是零向量（即，0+0i），一个复数z的逆就是$-z$，数1和i完美地构成这个空间的基。这是一个向量空间，有这样的附加特性：两个向量也就是两个复数可以相乘得到另一个复数。所以$\mathbb{C}$不只是一个向量空间，它还是一个代数。

把一个向量空间转变成一个代数不是一件很容易的事情，就像我们下一章的主角所研究的那样。例如，在前文中我介绍的那个六维多项式表达式的空间在通常的多项式乘法之下就不是一个代数。这就是为什么有用的代数往往很有名，因为它们稀有。为了创建一个代数，必须放宽某些规则，大多数情况是交换律，即$a\times b=b\times a$。通常还必须放宽结合律，即$a\times(b\times c)=(a\times b)\times c$。

如同把两个向量相乘一样，如果还想做除法，那么就必须更大幅度地限制你的选项。除非想放宽结合律，或者允许向量空间有无限维度，或者允许标量是比普通的数值奇怪得多的东西，否则就把这些选项一路缩小到了$\mathbb{R}$、$\mathbb{C}$以及四元数域。

终于引到四元数了！当当当当，威廉·哈密尔顿先生登场！

第8章

向第四维的跳跃

婉转点儿说，数学小说之类的文学作品数量并不多。在这类少见的作品中有一部很畅销——艾德温·A·艾勃特的《二维国》(*Flatland*)，初版于1884年，直到今天还一直在发行。

《二维国》是以一个自称“正方形”的生物的口吻讲述的故事。事实上，它是一个生活在二维世界“二维国”里的正方形。二维国还生活着各种其他生物，也都是按不同规则形成的二维形状：三角形［包括等腰三角形（有两个相等的边）和等边三角形（三个边全相等）］、正方形、五边形、六边形等。这个国度的生物也分三六九等，拥有的边越多地位就越高，圆是它们的老大。女性只是直线线段，遭受各种社会限制和歧视。

《二维国》的前半部分描述二维国和其社会的各个阶层。作者用大部分篇幅介绍了如何决定一个新生物在社会中的等级。因为二维国人的视网膜是一维的（就像你的视网膜是二维的一样），用这样的视觉看到的对象就是直线段，只有通过触摸来确定陌生对象的实际形状及它的等级。因此，介绍的通用形式是：“我请你摸一摸×××阁下。”

在这本书的后半部分，一个正方形视察了其他世界。在梦里他参观了线国，一个一维的地方，作者用了11页的篇幅介绍这部分内容。因为一维线国人永远无法穿越其两端的邻居，所以这一物种的繁殖就成了问题，艾勃特用极大智慧和才干解决了这一问题。

然后，正方形醒来，回到了自己的世界——二维国，不久之后，他将在这里接受来自三维的一个生物的造访：球。球以一种不稳定的方式伸伸缩缩地游荡到二维国，作为一个不可思议地伸张和收缩的圆来见正方形。这个球与正方形进行了一场哲学讨论，要点是要把他介绍给点国——一个零维空间，那里只居住着一种生物，“这种生物是它自己，它自己就是全部，这个零维空间什么也没有。然而，要留意他那十足的自满，从而吸取教训。他的这种自鸣得意让人厌恶，完全是无知的表现，所追求的就是盲目无知的幸福。”我想我们都遇到过这样的生物。

《二维国》发表时，艾勃特还不到46岁。1911年的大英百科全书中对他的描述有240个词条，把他描述成“英国校长和神学家”，但没有提《二维国》。艾勃特实际上是有进行思想改革倾向的男校校长，这是受到其天主教的个人观点和对当时维多利亚社会习俗的怀疑而促成的。《二维国》间接而温和地讽刺了社会。

自它出版120多年来，《二维国》引起了不计其数的广大读者的注意，激发了他们的想象力。事实上，后续的一系列书籍和故事都是以它为蓝本的。戴奥尼斯·伯格的《球国》和伊安·斯图尔特的《二维国内外》都是对艾勃特原创思想进行深加工的结果，这一点值得我们注意。1984年杜特尼在他的著作《平面世界》（*The Planiverse*）[1]中深入透彻地研究了二维世界的物理学、化学和生理学，对这些素材艾勃特明白但却不太相信。文学水准稍低但有一定吸引力的作品就是鲁迪·卢克的小说《二维国里的信息》（*Message Found in a Copy of Flatland*）。这本书的主人公在伦敦一位巴基斯坦居民的地下室中偶然发现了二维国。他最后吃掉了二维国人，这些人的“味道有些像水分很大的烟熏鲑鱼”。[2]

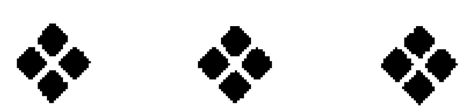
零维空间、一维空间、二维空间和三维空间，为什么就到此停止了呢？也许大多非数学家们都是从韦尔斯的1895年的小说《时光机器》中第一次听到第四维的，这本小说的主人公是这样说的。

> 我们的数学家研究的空间据说有三维，在这个空间中你可以指出长度、宽度和厚度，这些总可以由三个平面决定，每个平面都与其他两个平面成直角。但是一些哲学人士曾经问为什么要特地考虑三维？为什么没有另一个与这三个平面成直角的方向呢？他们甚至去尝试着构建四维几何。

一种深奥的理论出现后，是需要相当长时间才能渗入到受欢迎的文学作品中的，在今天也是如此。[3]在19世纪80年代到90年代出现的文学作品中提到也许存在也许不存在的空间维数，那么我们可以肯定地说，这样的问题当时职业数学家一定已经研究了几十年了。

事实确实如此。整个19世纪第二个25年间，各地的数学家开始产生关于空间维度的想法。在19世纪第三个25年间，这些零星的想法汇集到了一起，才使得伟大的德国数学家克莱因（1849—1925）做出了下面富有远见的发现："大约1870年，n维空间的概念已变成了年轻一代优秀数学家的必备知识。"

这些想法是从哪里来的呢？在19世纪初期它们还没有踪影，但到了19世纪末它们已广为人知，并在流行小说中出现了。又是谁第一个想到它们的呢？它们为什么又会在这个特殊的时期出现呢？

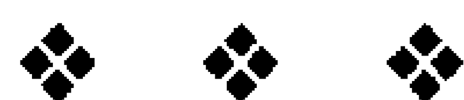


19世纪初的几十年间，成熟的复数概念已经作为普通知识扎根于数学家的大脑。如我在本书开始所描述的那样，数的现代概念已经在数学家的大脑中建立了起来。（但我前面用到的中空字母记号是在20世纪后期才开始使用的。）

而且，把实数和复数分别表示成沿直线上的点和遍布平面上的点的这种通用图表示已众所周知。复数巨大的能量以及它们在解决各种不同领域的数学问题中的用途也得到了广泛的认可。一旦所有这一切被人们所接受，自然就会引发下面的问题。

如果从一维的实数到二维的复数的变更会给我们带来这样巨大的力量和见识，那么为什么要止步呢？难道就没有正等待着被发现的其他种类的数，比如说超复数，它们本身就是三维的吗？还有，难道那些数就不能让我们进一步认识数学吗？

自从18世纪末开始，这一问题已经萦绕于几位数学家的脑海中了，这其中当然包括高斯，只是没有什么了不起的结果出现。大约在1830年，哈密尔顿也开始考虑这个问题了。



哈密尔顿[4]的人生履历其实是灰色的。这倒不是说他的生活境遇是多么不幸，比如受到战争、财产或者疾病等的困扰；也不是说他患有什么实质性的精神疾病，比如说长期的忧郁；更不是说他在事业上没有遇到伯乐或遭受到了挫败，他在十几岁时就已很有名了。我的意思是，哈密尔顿的人生走的是下坡路。当他还是一个孩子的时候，他就是一个神童；当他是一个青年人时，他只是一个天才而已；进入中年，他只能算一个比较聪明的人；进入晚年，他则成了一个令人讨厌的酒鬼。

哈密尔顿的数学才能是不容置疑的。他拥有深刻的数学洞察力，而且工作非常努力，因此使这一洞察力发挥到了极致，攻克了他那个时代最难的问题。直到今天，数学家们还是尊重他，以崇敬之情纪念他。

哈密尔顿出生于爱尔兰都柏林，其双亲是苏格兰人，所以人们说他既是爱尔兰人又是苏格兰人，但是他说自己是爱尔兰人。他是一个神童，少时一直在学习各种语言，到13岁时就宣称自己掌握了13种语言[5]，不再学这个了。1823年秋天，他进入都柏林三一学院，不久就脱颖而出成为著名的学者。大学第一年结束的时候，他遇到了凯瑟琳·迪士尼[6]并爱上了她。但是凯瑟琳的家人得知他们的关系后马上就把她嫁给了一个条件更合适的人。这次失恋毁了哈密尔顿的人生，可能同样也毁了凯瑟琳的人生。在1833年，哈密尔顿草率地与一位多病且精神不太正常的女人结了婚，余生都生活在这样病态的家庭中。

哈密尔顿在十几岁的时候就开始接触数学，很快就掌握了这门科学，并以在科学和古典文学课上取得最高分的优异成绩从三一学院毕业，这是一项前所未有的成就。在他毕业那年，他提出了“特征函数”，这是哈密尔顿算子的最原始概念，是现代量子理论的基础。

1827年，哈密尔顿被聘为三一学院天文学教授。像他那个时代其他年轻数学家一样，哈密尔顿被复数的精妙和力量所打动。1833年，他发表了一篇论文，用代数的方法看待复数系，这就是我们今天所说的“一个代数”。对于复数$(a+bi)$，哈密尔顿写的是(a, b)。复数的乘法规则变成：

$$(a,b)\times(c,d)=(ac-bd,ad+bc)$$

而且作为−1的平方根的i也可以表示成：

$$(0,1)\times(0,1)=(-1,0)$$

这看起来没什么了不起，那是因为从哈密尔顿时代至今数学已发展了170年。事实上，它完成了人们对复数的认识的变迁，是从19世纪初大多数数学家所定位的算术和分析的领域到代数领域的变迁。简而言之，它是向更高层次的抽象化和一般化的变迁。

从1835年开始，哈密尔顿开始了8年的探索，开发类似的三元组代数。因为从一维的实数到二维的复数就让数学有了这么大的发展，再增加一个维度难道不会有什么成果吗？

不过，事实证明从数值的三元组得到一个适当的代数是非常困难的。当然，如果你的条件足够宽松，那么你也能取得一定成果。然而，哈密尔顿知道他的三元组是与复数一样有用的东西，它们的加法和乘法必须满足相当严格的条件。例如，它们需要满足分配律，即，如果T_1、T_2和T_3是任意三元组，那么下面的等式成立：

$$T_1\times(T_2+T_3)=T_1\times T_2+T_1\times T_3$$

像复数一样，它们还需要满足模律。三元组(a, b, c)的模是$\sqrt{a^2+b^2+c^2}$。这个定律的文字描述是：如果把两个三元组相乘，那么其结果的模等于它们各自的模的积。

哈密尔顿一直冥思苦想这一问题，即把他的三元组转化成可运算的代数式，这一问题困扰了他8年。在这期间，他有了3个孩子，而且还获得了爵士的称号。

这一数学难题的解决得益于哈密尔顿那瞬间即逝的灵感。晚年他在给二儿子亚奇伯德·亨利的一封信里讲了这样一个故事。

> 1843年10月初，每天早晨我下楼吃早餐的时候，你的弟弟威廉·埃德温和你总是问我："爸爸，你能把三元组相乘吗？"而我总是悲伤地摇头，不情愿地回答说："不，我只能把它们相加和相减。"
>
> 但是，在10月16日，正好是星期一，爱尔兰皇家学会的议会日，我要去主持会议，你妈妈陪我行走在皇宫护城河边。尽管她不时跟我说着话，但此时一个想法突然闪现在我的脑海中，它最终给出了一个结果，当时无需多说什么，我立刻意识到它的重要性。似乎一条电路合上了，一道火花突然一闪，预示着在未来几年里我有了明确的思想和研究方向，如果我能活得足够长去展示这一发现，我要自己完成剩余的工作。我无法克制这一冲动，应该说是很不理智地立即拿出刀来，把下面的基本公式刻在布鲁穆桥的一块石头上，这是一个用符号i、j、k表示的基本公式：
>
> $$i^2=j^2=k^2=ijk=-1$$
>
> 这个公式包含了这个问题的解，当然这一题字将永远保存下去。

星期一在布鲁穆桥[7]上哈密尔顿突现的灵感，是三元组不能形成一个有用的代数，而四元组则可以。在从一维实数a到二维复数a+bi后，下一步不是到三维超复数(a+bi+cj)，而是跳到四维数(a+bi+cj+dk)。依靠某些简单的i、j、k的相乘规则，就是哈密尔顿刻在布鲁穆桥上的那些规则，

就能把这些四元组做成一个代数。范德瓦尔登在他的著作《代数历史》一书中把这一灵感称为“向第四维的跳跃”。

为了使这一新代数有效，他必须违反算术的一个基本规则——$a\times b=b\times a$。在第7章介绍柯西的置换合成时，我已经提到过这就是交换性。$a\times b=b\times a$这一规则就是交换律。这是我们大家在普通算术中所熟悉的，7乘以3的结果与3乘以7的结果相同。同样交换律也适用于复数：(2−5i)与(−1+8i)相乘的结果和(−1+8i)与(2−5i)相乘的结果一样（二者都等于38+21i）。

然而，只有放弃交换律，四元数才能像代数一样运算，四维向量空间中的两个向量相乘才有意义。例如，在四元数领域，i×j不等于j×i，而是等于它的相反数。事实上：

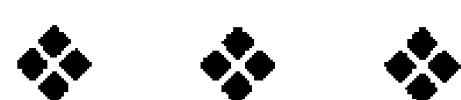
$$
\begin{array}{ll}
jk = i & kj = -i \\
ki = j & ik = -j \\
ij = k & ji = -k
\end{array}
$$

正是由于这些突破性的规则，四元数的代数意义才更加突出，并使哈密尔顿的洞察力的闪现成为数学历史中最具启示的一种。贯穿数值系统的所有革命，从自然数到远古时代的小数，到无理数和负数，再到复数和使18世纪及19世纪初伟大的数学家们绞尽脑汁的模算术，交换律总成立。现在，有了一种似乎可以被认为是一种数值的新系统，它不再满足交换律。借用现代术语，四元数是第一个非交换可除代数[8]。

每一个成年人都知道，当你打破了一个规则，再打破其他规则就容易多了。就如在日常生活中一样，代数的发现也是如此。事实上，哈密尔顿在1843年10月17日给他的一位数学上的知己约翰·格拉夫的信中描述了四元数。到了11月，格拉夫已经发现了一个八维代数，一种后来被称为八元数[9]的数系。然而，为了使其可运算，格拉夫不得不放弃另一个算术规则，即乘法的结合律，也就是$(a\times b)\times c=a\times(b\times c)$。

应该记住，这正是尼古莱·罗巴切夫斯基和雅诺·什鲍耶的非欧几

何（即“弯曲空间”）出现的时期。（在第13章中我将更详细介绍这一内容。）认为算术和几何法则是人类思维与生俱来的永恒特征这一康德思想已经过时。诸如交换律和结合律这样的基本定律都可以置之不理，那么还有什么东西同样可以置之不理呢？既然需要四维才能使四元数可行，有谁能说这个世界实际上不是四维的呢？[10]或者，也许在某个地方真有生活在二维国中的生物？



哈密尔顿的深刻洞察力只是那一时期前后出现的众多四维思想中的一个。在19世纪40年代，同四维一样，五维、六维以及*n*维也仍然悬而未决。如果说19世纪90年代是紫红色的十年①，那么19世纪40年代就是多维的十年，至少其间的数学家经过了这样的十年。

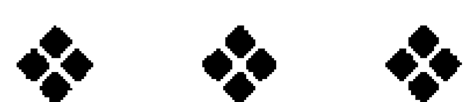
事实上，同是在1843年，我们将在下一章介绍的英国代数学家凯莱发表了一篇论文，题目是《*n*维几何新篇章》。正如这篇论文的题目所展示的那样，凯莱要说的是几何，但是他使用了齐次坐标系（我将在介绍代数几何时讲述齐次坐标系），这给这篇论文增添了强烈的代数色彩。

事实上，齐次坐标系是由德国天文学家和数学家奥古斯特·莫比乌斯想出来的，几年之前此人就这一课题出版了一本经典著作《重心演算》。莫比乌斯几乎在那时就已经知道通过四维旋转可以把一个不规则的三维固体转换成它的镜像。当时是1827年，可能实际是数学思想中最初的四维线索。

在这里提到莫比乌斯就不得不说说德国。虽然这些年是英国代数学家的天下，但是英吉利海峡对面的欧洲大陆的德国也已经涌现出很多天才，他们使德国站到数学世界的前列，并在19世纪后期和20世纪初期一直保持着这样的领先势头。

---

① 19世纪90年代，一大批以唯美主义者和颓废派自居的艺术家特立独行，把19世纪末的欧洲文坛搅闹得声色绘呈，形成了当时欧洲文坛一大景观，被称为“紫红色的十年”（Mauve Decade）。——编者注



这一时期，在普鲁士的斯德丁（现在波兰的什切青）有一位名叫赫尔曼·甘特·格拉斯曼的中学校长。1843年时他34岁，从1831年开始他一直在这个高中任教，一辈子都在教书育人。他的数学是自学的，他在大学时学的是神学和哲学。他40岁结婚，有11个孩子。

哈密尔顿发现四维向量空间的第二年，格拉斯曼出版了一本名字很长的书——《……线性扩张论》，通常我们叫作《扩张论》。在这本书中，格拉斯曼研究了80年后才出现的向量空间理论。他定义了一系列概念，诸如线性相关、线性无关、维度、基、子空间和投影。事实上，不仅如此，他还给出了向量乘积的方法和代表基的变换方法，因此也就用比哈密尔顿的四元数更通用的方法发明了“一个代数”的概念。所有这一切都是按很强的代数风格完成的，强调了这些新数学对象完全抽象的属性，仅仅是作为这些代数对象的应用引入了几何的思想。

但是，格拉斯曼的书几乎没有得到关注。只有格拉斯曼自己写了一篇评论。事实上，格拉斯曼就如阿贝尔、拉菲尼、伽罗华等这些非常悲哀的数学家一样，他们的价值不会被同代人接受。另外，也有他自己的问题。他的著作《扩张论》的写作风格按19世初的习惯充满了形而上学的色彩，令人难以理解。莫比乌斯对它的描述是“无法阅读”，尽管他尝试着帮助格拉斯曼，并于1847年写了一篇评论赞扬了格拉斯曼的想法。格拉斯曼全力推销这本书，但是他运气不好，没人理睬。

法国数学家圣维南1845年发表了一篇关于向量空间的论文。这是《扩张论》出版后的第二年。尽管圣维南与格拉斯曼不认识，但他的这篇论文展示了与格拉斯曼类似的想法。读了这篇论文后，格拉斯曼把《扩张论》中的相关段落寄给了圣维南。但他不知道圣维南的地址，只好通过法国科学院的柯西转交，并请求柯西尽快送到。然而，柯西没有这样做，并且于6年后发表了一篇论文，这篇论文非常有可能源自格拉斯曼的著作。格拉斯曼向科学院提出了抗议。尽管科学院成立了三人委员会来调

查是否发生了剽窃，但最终没有得出任何结论……

《扩张论》也并非完全不可读。哈密尔顿就于1852年看了这本书，并在其著作《四元数讲义》的引言中对格拉斯曼做了短评。他赞扬《扩张论》是“原创和杰出的”，但是强调他自己的方法不同于格拉斯曼的方法。就这样，格拉斯曼的著作出版9年后，有两位认真严谨的科学家注意到了它：莫比乌斯和哈密尔顿。

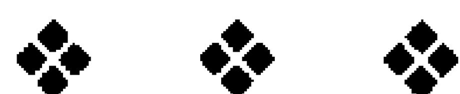
格拉斯曼尝试着重写《扩张论》，使它更容易被人接受，并于1862年自己出资出版了新版本，共有300页。这一新版本的前言中有如下一段话，我觉得它相当感人。

> 我始终坚信我在此科学上所付出的劳动不会白费，它耗尽了我生命中最重要的阶段，让我付出了超常的努力。我当然知道我给出的这门科学的形式还不完善，它一定是不完善的。但是，我知道而且有义务在此声明（可能有人会认为我很狂妄），即使这一成果再过十七年或更长时间还不被使用，也没有真正融入到科学的发展之中，它冲出遗忘的尘埃现身的时候也一定会到来，现在沉睡着的思想结出硕果的那一天一定会到来。我知道，如果我今天还不能（如我至今徒劳地期望那样）把学者们吸引到我的周围，用这些思想帮助他们成果累累，促使其进步，丰富其学识，那么这种思想在将来一定会重生，或许以另一种新形式，与时代发展水乳交融。因为真理是永恒不灭的。

1862年版的境遇比1844年的稍好些。然而，由于理想的破灭，格拉斯曼不再研究数学，而是投身于梵语研究。他出版了一本大部头的翻译著作，把梵语的经典著作《梨俱吠陀》翻译成德语，附加上很长的注释，总共有3000页。因为这一成就，他获得蒂宾根大学授予的荣誉博士称号。

直接以格拉斯曼的工作为基础的第一个重要的数学进步发生在1878年，这时他已经去世一年了。当时英国数学家克利福德发表了一篇题为

《格拉斯曼扩展代数的应用》的论文。克利福德利用格拉斯曼的想法把哈密尔顿的四元数一般化成为n维代数族。后来证明这些克利福德代数可以用于20世纪的理论物理。n维空间内的旋转（即旋子的现代理论）就来自于它们。



因此，19世纪40年代诞生了两个全新的数学对象，只是它们那时还没有被人理解，其发明者也没有给它们起这样的名字：向量空间和代数。即使在发展的初级阶段，这二者也为数学研究创造了大量机会。

对于实际应用也是如此。当时是电气时代的初期。哈密尔顿伟大的灵感闪现距离迈克尔·法拉第发现电磁感应相隔了12年。当时法拉第已经是52岁了，但仍然很活跃。四元数发现两年后的1845年，法拉第提出了电磁场的概念。他完全用虚构的术语（如“力线”等）来描述它，因为当时的数学知识不足以让他把想法表述得更加严谨。他的后辈，著名的麦克斯韦，具有丰富的数学知识，他发现向量正是他们表示这些新认识所需要的东西。

事实上不难想到，在那个时代对于这种新的奇妙的电气科学（各种强度的电流流向各个方向）的兴趣是催生向量思想的主要动力之一[11]。不完全是物理学家发现在工作中使用向量很容易，事实上从19世纪末开始存在三个思想学派。

向量思想的第一个学派源自哈密尔顿，他是在现代科学中使用向量和标量的第一人。哈密尔顿认为四元数（$a+bi+cj+dk$）是由标量部分a和向量部分（$bi+cj+dk$）组成的，而且发明了建立在这一体系之上处理向量的方法，即把向量和标量一起捆包在四元数包里。

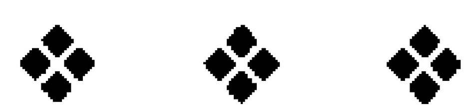
第二个学派是由美国人吉布斯和英国人赫维赛德于19世纪80年代建立的，他们把四元数的组成成份向量和标量分开，把它们作为独立的实体进行处理，并创建了现代向量分析。最终的结果实质上是格拉斯曼体系，尽管吉布斯证明在他看到格拉斯曼的著作之前，他的想法已经形成，

而赫维赛德似乎根本没有看到过格拉斯曼的著作。吉布斯和赫维赛德是物理学家，不是数学家，他们有着纯数学家们谴责的那种自命不凡的实验主义观念。他们只是想要对他们有用的代数，如果这意味着从哈密尔顿四元数那里获得了想要的东西，他们不会为这样的做法感到内疚。

第三个学派以英国科学家开尔文（威廉·托马斯）为代表，他完全避开这些全新的数学，采用古老的笛卡儿坐标（x, y, z）进行研究工作。这种符合英国人性格的保守方法至少在英国使用了很长时间，对那些文化修养不高的人很有吸引力。20世纪60年代我从一位上了年纪的中学校长身上体会到了这种社会力量的强大，他是开尔文阵营中的铁杆分子，并声称向量“只是一时的时尚”。

关于这些体系的优势之争引发了19世纪90年代有点可笑的伟大四元数大战，保罗·纳欣在他1988年的赫维赛德自传中对这场混战做了很详细的记述（第9章）。格拉斯曼才是这场混战最后的胜利者，或者如你所说，是向量的胜利者。把四元数放到一旁，利用笛卡儿坐标进行数学物理研究的开尔文方法反倒显得有点奇怪和笨重。

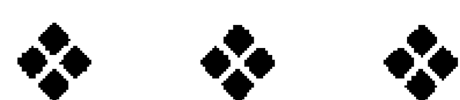
因此，四元组代数从没有实现哈密尔顿对它寄予的厚望，而只是起了间接作用。四元数辜负了哈密尔顿的期望，辜负了他生命最后20年的辛勤工作，它没能够打开一片新数学天地。事实证明，四元数形式理论成为数学的一坛死水，只是由于在几个深奥的高等代数领域需要这一理论的关系，才作为群论或矩阵理论课程的简要补充教授给本科生[12]。



在四元组代数年代，n维空间的研究走到了其他方向。19世纪50年代初，瑞士数学家路德维希·施拉夫创建了四维或更高维空间内的“多面体”几何。所谓多面体就是有平坦侧面的图形，它们与二维多边形和三维多面体类似。施拉夫关于这一话题的论文的英文版和法文版分别发表于1855年和1858年，比格拉斯曼的《扩张论》下场还悲惨，几乎完全没人理睬。在他死后的1895年，他的著作才得见天日。这一工作应该属于

几何范畴而不是代数范畴。

另一条发展路线起源于黎曼于1854年发表的伟大的博士学位资格讲演，其题目是《几何基础之假设》。黎曼重新拾起高斯已经放下的某些想法，改变了人们对曲线和表面几何的整体观点。例如，人们不再把弯曲的二维表面看成被嵌入到平坦的三维空间中的对象，而是通过不能离开该表面的生物从内部研究这个表面。他的"内蕴几何"可以容易且显然地推广到任意维，导致了现代微分几何、张量微积分以及广义相对论的诞生。当然这一话题不是纯代数话题（在第13章讨论现代代数几何时，我会再讨论这一话题）。



抽象向量空间和代数空间（也就是按着某种方式可以进行向量乘法的空间）的理论最终发展成为一个范围广阔的领域——线性代数。一旦从向量乘积的交换律和结合律的束缚中解放出来，各种奇怪的东西就会出现，必须将它们合并到一个更一般的理论之中。

例如，有一些代数的零向量有因子！实际上，当哈密尔顿尝试将他的四元数一般化，使得四元数a+bi+cj+dk的系数a、b、c、d不再只是实数（如他最初所认为的那样）而是复数时，他自己也意识到了这一点。例如，在复数域上，可以做下面的因式分解：

$$x^2+1=\left(x+\sqrt{-1}\right)\left(x-\sqrt{-1}\right)$$

（在这里我写了$\sqrt{-1}$，避免与哈密尔顿四元数中的i混淆，四元数中的i与复数的i不是一回事。）如果用哈密尔顿的j替换x，就得到

$$\mathrm{j}^2+1=\left(\mathrm{j}+\sqrt{-1}\right)\left(\mathrm{j}-\sqrt{-1}\right)$$

但是，根据哈密尔顿的定义，$\mathrm{j}^2=-1$，所以$(\mathrm{j}+\sqrt{-1})$和$(\mathrm{j}-\sqrt{-1})$是零的因子。在现代代数中这种情况并不是唯一的。考虑下一章我要介绍的矩阵乘法，当把两个非零矩阵相乘时，可能会得到一个零矩阵的结果。然而，这一结果的确表明抽象代数的研究已经远远偏离了实数和复数范畴。

列举所有代数并对其分类是一件很有趣的练习。你的结果取决于你愿意放弃什么。最有限的情况是实数域ℝ上不允许除0的可交换、可结合有限维代数。正好有两个这样的代数：ℝ和ℂ，这是1864年由卡尔·威尔斯特勒斯证明的。依次放宽规则，允许不同的基础域成为标量，允许零因子等诸多项存在，就会得到越来越多的代数，而且这些代数具有越来越多的奇异性质。美国数学家本杰明·皮尔斯于1870年沿着上面这一路线进行了一次著名的分类。

苏格兰代数学家韦德伯恩在1908年的一篇题为《论超复数》的论文中把代数推广到了一般化的新层次，允许标量是任意域……但是，我已经提到了域和矩阵等话题，到现在都还没有详细描述它们。特别是源于古老的中国的矩阵，单单这一话题就需要一章的篇幅来介绍。

注 解

[1] 《平面世界》是我非常喜欢的书。该书妙趣横生，可作为锻炼想象力的读物。例如，二维生物如何敲门？如果它有一条从身体一头流向另一头的营养导管，那么怎么才不至于成为两个分离的部分？

[2] 这个故事取自于《数学航海》——1987年由卢克编写的数学科幻故事文选。在这一集中有23个故事，我认为其中一半以上都是在展示四维的思想。

[3] 20世纪90年代一些不负责任的小说家认为韦尔斯于1895年所写的小说《时光机器》如果不加上海森堡的不确定性原理是不完整的，真可笑，这一原理1927年才第一次提出来！

[4] 关于哈密尔顿的文学作品有很多，全方位的传记至少就有3本。我主要是依据1980年由托马斯·汉金斯写的传记，同时又参考了一些数学杂志、数学教科书和网站。

[5] 在19世纪初期的英国和美国，这样的说法好像很普遍。《圣经在西班牙》的作者乔治·博罗比哈密尔顿大两岁，人们也猜测他掌握了很多种语言。安妮·里德勒博士对博罗在语言方面的技能做了研究，确定他有51种语言和方言的阅读能力。在这一研究中，里德勒博士提到了1793年出生的《兄弟乔纳森》的作者、美国作家约翰·尼尔。尼尔曾经声称："我在两三个月内就熟悉了法语、西班牙语、意大利语、葡萄牙语、德语、瑞典语、丹麦语，除此之外还复习了希伯来语、拉丁语、希腊语和撒克逊语等。"诗人朗费罗比哈密尔顿小一岁半，在他仅有19

岁的时候，就被任命为鲍登学院的现代语教授，证明他的确掌握了很多现代语言。他自学了法语、西班牙语、意大利语和德语，达到了熟练阅读的水平，我们主观认为在1826年到1829年之间，他分别用了9、9、12和6个月时间就掌握了这些语言。

[6] 人们很自然想问凯瑟琳是否与华特·迪士尼（迪士尼品牌创始人）有关系。也许有，但是我始终没有找到任何证据。迪士尼是一个有很多爱尔兰分支的古老家族。华特出身于华特·里亚斯·迪士尼家族，大约1803年出生于爱尔兰。说华特是西班牙人的私生子并被迪士尼家收养的故事只是民间传说。

[7] 布鲁穆桥在都柏林工业区，大约距市中心西北5公里。

[8] 我们已经知道了，早在1820年高斯就构想了一个非结合性的代数，只是不愿费心去发表。要想做得比伟人好，你不得不比他付出更多才行。

[9] 1845年凯莱独立发现了八元数，因此有时候也称其为凯莱数。

[10] 这个世界是四维的吗？在任何简单的几何意义下，这个世界都不是四维的。这里没有“第四个方向”，尽最大努力去想象，你也许可以想象出四维世界。如果你真进入了四维世界，也许会立刻崩溃，因为即使是平方反比律这样最简单的物理定律，若要把它搬到四维欧几里得空间，都会导致非常悲惨的后果。当然，现代物理学的时空用四维几何描述很方便，但本质上那种几何根本就不是欧式几何，所以你应该从头脑中彻底清除带着普通的欧式几何到四维旅行的想法。然而，人类的想象力是一件非常奇妙的东西。已故的考克斯特在他的著作《正规多面体》中提到他的朋友约翰·弗林德斯·皮特里说：“在精神高度集中的时候，他能够通过‘可视化四维图形’来回答关于它们的复杂问题。”

[11] 在欧洲大陆，情况却完全相反，在那里法拉第所钟爱的“力线”没有更古老的“超距作用”的思想受欢迎。翻阅一下当时的数学文献，包括德国的数学文献，你会看到这些作者们还不知道表面及空间的有向流等思想。

[12] 四元数在量子理论中有一些应用。引用一位给予我帮助的物理学家朋友的一些笔记：“有趣的是，如果一个人要用四元数来公式化旋转运动学，那么其结果的7×7的协方差矩阵（里卡蒂方程的解）是唯一的，因为4参数欧拉对称参数是线性相关的。”正是如此。康威和史密斯于2002年合著《四元数和八元数》一书对哈密尔顿的想法给出了非常全面的概括，但其数学层次很高。印地安那大学教授安德鲁·汉森有一本名为《可视化中的四元数》的著作，与我的这本书差不多在同一时间（即2006年初）出版。我没有看过那本书，但是其中很多内容是关于四元数在计算机动画方面的应用的。

第9章

项的长方形排列

下面要说的其实是一个文字问题。我发现，如果把三种不同的谷物想成颜色各不相同，例如红、蓝和绿，那么这个问题就很容易想象。

> 问题。有三种谷物。如果是第一种3袋、第二种2袋、第三种1袋，那么总重量为39个重量单位。若第一种2袋、第二种3袋、第三种1袋，总重量就是34个重量单位。若第一种1袋、第二种2袋、第三种3袋，总重量就是26个重量单位。请问，每种谷物一袋所含谷物是多少个重量单位？

假设红色谷物一袋的重量是x，蓝色谷物一袋的重量是y，绿色谷物一袋的重量是z。那么必须求解下面关于x、y、z的联立线性方程组：

$$3x+2y+z=39$$
$$2x+3y+z=34$$
$$x+2y+3z=26$$

很容易就可求得，解是x=37/4、y=17/4、z=11/4。

无论是丢番图还是古巴比伦的数学家，都不难得到这个问题的答案。这个问题之所以在数学历史中占据如此显著的位置，是因为提出这个问题的作者开发了解决这一问题及任意类似的多个未知数问题的系统方法，直到今天，刚开始学习矩阵代数的学生们还要学习这种方法。而所有这一切竟然发生在2000年之前！

我们不知道这位数学家的名字。他可能是《九章算术》的作者或者编辑。这是一本有246个丈量和计算问题的问题集，是古代中国数学史上当之无愧的最具影响力的著作。它对中世纪印度、波斯、穆斯林以及欧洲的数学的影响程度存在争议，但是这本书的译本的确从公元后不久以来就传遍了东亚，我们可以通过它了解中世纪穿越欧亚大陆的贸易和文化往来，如果某些西亚和西方数学没有从它那里得到灵感，那才让人大跌眼镜。

根据这本书自身包含的一些证据和后来版本的编辑的一些评论，我们可以确定《九章算术》原创于汉朝（公元前202年到公元9年）。汉朝是中国历史上一个伟大的新纪元，是国土面积覆盖了现代中国大部分地区的第一个朝代，是由汉族人牢牢统治的朝代。

实际上，中国文化版图早在公元前211年的秦代就已经被著名的令人闻风丧胆的“始皇帝”统一了。然而在这个暴君死去的11年后，秦的政治体系迅速瓦解。随后开始了多年的内战（这为中国文学作品、戏剧和歌剧提供了诸多素材），最后刘邦击败了对手登上了皇帝宝座，并于公元前202年建立了汉朝。

秦始皇臭名昭著的行为之一就是烧毁书籍。根据一位名叫商鞅的哲学家的严格的集权主义教条，秦始皇命令把所有持可疑观点的书籍都收集到权威部门烧毁。幸好，在古代中国，学习主要是靠死记硬背，[1]所以秦政权瓦解之后，头脑中仍然记得那些被焚毁的文献的学者们就能够重新整理这些书籍。可能就是在这个时候，《九章算术》才得以最终成形，把来自各方面记忆的文献汇总起来。也有可能不是在秦始皇焚书后学者们重新整理的：这位暴君“赦免”了那些关于农业和其他实践课题的书籍，因而如果《九章算术》早于汉代，那么它很可能没有被烧毁。

总之，汉代初期是中国数学的创造期。和平促使贸易发展，而贸易需要计算技术。开始于秦的重量和丈量标准使得人们开始研究面积和体积计算。确立儒教作为国家信条需要一个精确的历法，这样才能在正确

的时间进行正确的观测。一种以通常的19年周期为基础的历法[2]应运而生。

《九章算术》很有可能是汇集当时各方面之创造的成果。这本书在公元1世纪就已存在，而且在随后的中国数学文化发展历程中承担了与欧洲欧几里得《几何原本》类似的角色。在其中的第8章，就有我所描述的谷物称重问题。

《九章算术》的作者是如何解决这个问题的呢？首先，把三个方程中的第二个方程乘以3（于是方程变成$6x+9y+3z=102$），然后用它减去第一个方程两次。类似地，把第三个方程乘以3（使得它变成$3x+6y+9z=78$），然后用它减去第一个方程一次。此时这三个方程变成：

$$3x+2y+z=39$$
$$5y+z=24$$
$$4y+8z=39$$

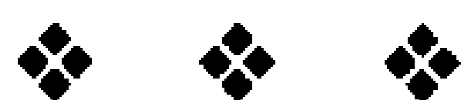
现在，把第三个方程乘以5（使之变成$20y+40z=195$），然后用它减去第二个方程四次。于是第三个方程简化成：

$$36z=99$$

从上面的方程可以得到$z=11/4$。把这一结果代入到第二个方程得出y的解，把z和y的值代入到第一个方程得到x。

可见，这是一个通用性很强的方法，它不仅可以用于有3个未知量的3个方程，也可以用于有4个未知量的4个方程、5个未知量的5个方程等。

这一方法就是我们现在所知道的高斯消元法。伟大的高斯在1803年到1809年对雅典娜星座做了几次观测，并计算了这个对象的轨道。这一计算中就要涉及求解6个未知量的6个线性联立方程。高斯就像我上面做的那样解决了这个问题，而《九章算术》的未知名作者在2000年前就做到了。



一旦手中有了好用的文字符号体系，那么很自然就有这样的疑问：

如果对下面这样的一般三次联立方程组尽量用高斯消元法，会得到x、y、z什么样的解？一般三次联立方程组为：

$$ax+by+cz=e$$
$$fx+gy+hz=k$$
$$lx+my+nz=q$$

它的解是：

$$x=\frac{bhq+ckm+egn-bkn-cgq-ehm}{agn+bhl+cfm-ahm-bfn-cgl}$$
$$y=\frac{ahq+ckl+efn-akn-cfq-ehl}{agn+bhl+cfm-ahm-bfn-cgl}$$
$$z=\frac{agq+bkl+efm-akm-bfq-egl}{agn+bhl+cfm-ahm-bfn-cgl}$$

就是这样复杂的事情让人对数学望而生畏，进而放弃。然而，如果你能坚持不懈，就能从中发现某种模式。例如，上面三个解的分母是一样的。

现在把注意力转向分母。这个分母的表达式是$(agn+bhl+cfm-ahm-bfn-cgl)$。注意$e$、$k$和$q$在这个表达式中根本没有出现，该表达式完全是由等号左边的系数构成的，即来自于下面的矩阵：

$$\begin{matrix} a & b & c \\ f & g & h \\ l & m & n \end{matrix}$$

下面要观察的是：分母中这6个项中的任意一个都不会在这个矩阵的同一行或同一列出现两次。

例如，观察ahm项。a取自第一行、第一列，就好像把第一行和第一列划出这个矩阵。下一个数h，它不能取自第一行和第一列，而必须取自其他行和列。于是从第二行和第三列取得h之后，这个行和列就被划掉。接着是m，除了从第三行和第二列选出m外没有其他选择。

不难证明，把上面的逻辑运用于3×3的矩阵，结果就是所有可能的6个项：对于2×2矩阵，会得到2个项；对于4×4的矩阵，就会得到24个项。

这个项数是我在第7章中介绍的阶层数：2！=2，3！=6，4！=24。对于有5个未知量的5个方程，其相应的数值是5！=120。

各项的符号更加麻烦。一半是正的，另一半是负的，但是如何确定哪个是正哪个是负呢？为什么*agn*项是正的，而*ahm*就是负的呢？仔细观察。

首先注意，我是按行取这些字母，再依顺序写出项的。比如项*ahm*，*a*是第一行，*h*是第二行，*m*是第三行。当然，也完全可以按这些字母所在列的顺序没有歧义地把项写出来。对于*a*、*h*和*m*的相应列分别是1、3和2。只要我坚持按行的顺序写系数，那么三元组(1, 3, 2)就是*ahm*的一种表示。

当然，这个三元组也是基础三元组(1, 2, 3)的一个置换，如果把基础三元组(1, 2, 3)中2的位置放3，3的位置放2，就可以得到上面的三元组。

现在，置换分两种：奇置换和偶置换。上面的置换是奇置换，这就是为什么*ahm*有负号的原因。另一方面，项*bhl*的表示是(2, 3, 1)，它是通过把基础三元组(1, 2, 3)中的1变成2、2变成3、3变成1而得到的，因此这个置换是偶置换，所以*bhl*有正号。

太美妙了，但是又如何知道一个置换是奇置换还是偶置换呢？下面就说如何知道。我将继续使用有3个未知量的3个方程的表达式，不过推广到4、5及任意多个方程上去也很容易。

*A*和*B*为1、2或3，且$A>B$的所有可能数值组合成$(A-B)$并相乘，得到$(3-2)\times(3-1)\times(2-1)$。这个积的值是2（它就是$2!\times1!$，所以很容易得到推广后的形式。如果处理的是有4个未知量的4个方程，这个积是$(4-3)\times(4-2)\times(4-1)\times(3-2)\times(3-1)\times(2-1)$，它的值是12，即3！×2！×1！）。但是这个值并不重要，重要的是它的符号，它的符号是正的。

现在，对前面表达式中的数值1、2、3实施某个置换。先尝试第一个置换，即*ahm*。我们重新排列基础三元组，把2变成3，3变成2。现在，这个表达式变成$(2-3)\times(2-1)\times(3-1)$，计算结果是−2，所以这是一个奇

置换。运用于置换*bhl*，把(3−2)×(3−1)×(2−1)变成(1−3)×(1−2)×(3−2)，这一表达式的计算结果是+2，这是一个偶置换。

奇偶置换的这一性质非常重要，显然与我在第7章中讨论的问题相关。下面是(1, 2, 3)的所有6种可能置换，它们的奇偶性用我刚才使用的方法可以计算出来。

$$
\begin{array}{ll}
(1,2,3) & (3-2)\times(3-1)\times(2-1)=2 \\
(2,3,1) & (1-3)\times(1-2)\times(3-2)=2 \\
(3,1,2) & (2-1)\times(2-3)\times(1-3)=2 \\
(1,3,2) & (2-3)\times(2-1)\times(3-1)=-2 \\
(3,2,1) & (1-2)\times(1-3)\times(2-3)=-2 \\
(2,1,3) & (3-1)\times(3-2)\times(1-2)=-2
\end{array}
$$

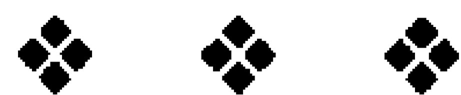
可见，它们中有一半是奇置换，另一半是偶置换。总是这样的。

所以，那个分母表达式中每一项的符号是由对应于系数表示的置换的符号决定的。终于找到规律了！

我刚才使用的分母表达式是*行列式*的一个例子。事实上，对于*x*、*y*和*z*，分子也是行列式。你可以尝试弄清楚每一个分子是哪一个3×3数组的行列式。行列式的研究最终导致矩阵的发现，现在这是代数中非常重要的课题。一个矩阵通常是数值的一个数组，例如像我上面所给出的3×3矩阵，但是，矩阵自身是作为一个对象来处理的。详情后述。

奇怪的是，现代代数课程是先介绍矩阵再介绍行列式，与发现它们的历史顺序相反：在矩阵之前很早人们就知道行列式了。

这样做的根本原因是行列式是一个数（对此我还会进行补述），而矩阵不是数，它是一种不同的东西，是*一个不同的数学对象*。本书现在研究的这一时期是19世纪初期和中期（尽管没有在本章讲述，但还有一段历史需要补充），当时代数对象正将自己从数和位置这样的传统数学中分离开来，开始自己的生活。



虽然几个世纪以来漫不经心地处理着线性联立方程组的数学家一定

在无意中多次发现了行列式类型的表达式，而且我提到过的一些代数学家，特别是卡尔达诺和笛卡儿等人，已经离发现这一实体很近了。但是，直到1683年才毫无争议地清晰地发现了行列式。这是数学历史上最著名的巧合事件之一，那一年两次发现了行列式。一次是在汉诺威王国，现在是德国的一部分；一次是在江户，就是今天的日本东京。

这位德国数学家当然是我们熟悉的一位数学家，他就是莱布尼茨，微积分的联合发明者之一（是与牛顿同时发现的），他还是哲学家、逻辑学家、宇宙学家、工程师以及法律和宗教的改革者，非常博学，如他的一位传记作家所说的那样，莱布尼茨是一位“整个脑力研究世界的市民”。[3] 莱布尼茨年轻时曾四处游历，之后，他70年生涯中的后40年都在为汉诺威大公服务。汉诺威是诸联邦州中最大的一个，在地图上所占据的位置现在属于德国。

莱布尼茨在1683年写给法国数学精英洛必达的一封信中说，如果两个未知数的3个方程的联立线性方程组

$$a+bx+cy=0$$
$$f+gx+hy=0$$
$$l+mx+ny=0$$

有解x和y，那么

$$agn+bhl+cfm=ahm+bfn+cgl$$

这就相当于说行列式（$agn+bhl+cfm-dhm-bfn-cgl$）一定等于零。莱布尼茨是完全正确的。尽管他没有创造行列式的完整理论，但是他的确清楚地知道在求解联立线性方程组的过程中它们的重要性，并掌握了掌管行列式的结构和行为的一些对称原则，就像我在前面陈述过的那些原则。

莱布尼茨的联合发现者是日本的关孝和，关于他的出生和死亡的具体时间我们都不太清楚，他大约于1642年或1644年出生在江户（今天的东京）。为《科学传记辞典》撰写关孝和的词条时，小堀宪是这样说的：“对关孝和生平的了解是不全面且间接的。”他的生父是一名武士，但是

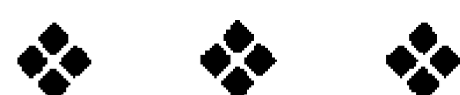
关家收养了他，所以他随了他们的姓。

当时，在强大政权的统治下，日本已经在几十年前就进入了国家统一和自信的时代，即江户时期，其中第一位也是最伟大的一位幕府将军是詹姆斯·克莱威尔的一本华美小说的主题。和平统一使得货币流通，因此对会计和审计员的需要量大增。关家的人都从事相关的工作，关孝和也走上了这条路，最终被提升负责全国的物资供给，还被晋升成为武士。据《科学传记辞典》记载，“1706年由于年龄过大无法胜任其职务，他被换到一个清闲的职位，并于两年后去世。”

由日本人写的第一本数学著作出版于1622年，就是森成吉的《论除法》。我们的主人公关孝和是森成吉的学孙，也就是说关孝和的老师（这个人名叫高原，关于此人我们几乎一无所知）是森成吉的学生。关孝和深受中国数学文献的影响，毫无疑问他知道《九章算术》。

当时的东亚尽管也发明了系数、未知量和未知量的幂的“文字”（这些“字母”实际上就是中国文字）符号体系，但是关孝和知道的不仅仅是东亚的这些方法。虽然他的方程解是数值解而不是严格的代数解，但是他的研究却很深刻，他几乎发明了微积分。今天我们所说的由杰克·伯努利于1713年引入欧洲数学的伯努利数，实际上关孝和在他发现的30年前就已经发现了[4]。

作为一名武士，人们希望关孝和行为谦虚，因此他出版的书籍不能用他的本名。这里还涉及数学教育相互竞争的各校之间相互保密的文化，非常像我之前所描述的16世纪意大利学校那样。我们知道关孝和的工作是通过他的学生出版的两本书，一本出版于1674年，另一本则在他死后的1709年出版。在这第二本书中出现了关孝和关于行列式的研究。他提炼并扩展了我之前所讲述的中国的行消元法，而且展示了其中的行列式部分。

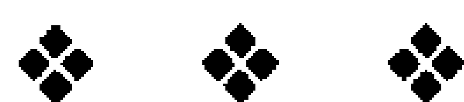


遗憾的是，无论是关孝和的工作还是莱布尼茨的工作都没有很多直

接成果。直到现代，在日本也似乎没有进一步的发展。在欧洲，当行列式再次被提出时，已经整整经历了一代人。随后，它突然开始流行，到了18世纪已经成为西方数学的主流。我在上文中已大致描述了利用行列式求解线性联立方程组的过程，这一过程被称为克莱姆法则。它第一次出现在1750年出版的名为《代数曲线分析导论》的书中。这本书的作者是加百列·克莱姆，一位瑞士数学家和工程师，他四处游历，与他那个时代的欧洲伟大的数学家都非常熟悉。在他的书中，克莱姆解决了寻找经过平面上任意n个点的简单代数曲线（即满足自身的x、y方程在等号的左边是多项式、右边是零的曲线）问题。他发现，任给5个点，可以找到一条符合它们的二维曲线，即有如下方程的曲线：

$$ax^2+2hxy+by^2+2gx+2fy+c=0$$

我将在后面代数几何知识中详细介绍这一类曲线。寻找实际给出的5个点的曲线方程需要求解一个有5个未知量的线性联立方程组。这不仅仅是一个抽象的问题。根据开普勒定律，行星在这样一条（非常近似的）二次曲线上运行，因此，5次观测行星位置就足以相当精确地确定它的轨道。[5]



行列式之间可以进行运算吗？例如，给定两个方形数组，如果我要按如下所示把它们相应的元素加起来得到一个新数组：

$$\begin{matrix} a & b \\ c & \mathrm{d} \end{matrix} + \begin{matrix} p & q \\ r & s \end{matrix} = \begin{matrix} a+q & b+q \\ c+r & \mathrm{d}+s \end{matrix}$$

那么这个新数组的行列式是前两个数组行列式的和吗？不是，左边两个数组的行列式加起来是$(ad-bc)+(ps-qr)$，而右边这个数组的行列式是$(a+p)\times(d+s)-(b+q)\times(c+r)$，它们不相等。

尽管行列式相加不成功，但它们相乘可以。让我把上面两个行列式相乘：$(ad-bc)\times(ps-qr)$等于$adps+bcqr-adqr-bcps$。对于所有我们感兴趣

的行列式都成立吗？事实正是如此。下面就是2×2数组的行列式：

$$\begin{matrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{matrix}$$

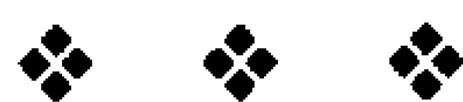
仔细观察上面的数组，就会看到它的4个元素中的每一个元素都是第一个数组的行元素（a、b或者c、d）和第二个数组的列元素（p、r或者q、s）经过简单算术得到的。例如，第二行、第一列上的元素来自于第一个数组的第二行和第二个数组的第一列。这不仅仅适用于2×2数组：如果把两个3×3数组的行列式相乘，就得到一个表达式，它是另一个3×3数组的行列式，这个“积数组”的第n行和第m列上的数值是通过合并第一个数组的第n行与第二个数组的第m列得到的，这一过程刚好与上面描述的过程类似。[6]

对一名19世纪初期的数学家来说，观察这个2×2乘积数组会令他想起另外一件事：高斯1801年的经典著作《算术研究》中的第159节。在那里，高斯提出了下面的问题。假设我对x和y的某个表达式做这样的替换，$x=au+bv$，$y=cu+dv$。换句话说，经过一个线性转换，把未知量从x和y变成u和v，x是一个u和v的线性表达式，y也同样。假设再一次做这样的替换，使用另外一对未知量w和z，经过另外两个线性转换$u=pw+qz$和$v=rw+sz$。

其综合效应是什么呢？在从未知量x和y到未知量w和z变化的过程中经过了中间未知量u和v，中间还有一次线性转换，其结果转换是什么呢？这很容易弄明白。这个综合效应就是这样的替换：$x=(ap+br)w+(aq+bs)z$，$y=(cp+dr)w+(cq+ds)z$。观察括号里的表达式！看起来行列相乘与线性转换有某种联系。

这些想法和结果的传播，使得建立清晰易懂的行列式理论只是时间问题。这一工作是由柯西完成的，1812年他在法国科学院宣读的一篇很长的论文中发表了这一成果。柯西对行列式及其对称性和操作法则等给出了完整而系统的描述。他还描述了我在这里给出的乘法法则，当然他

的法则比我给出的更通用。柯西1812年的论文被认为是现代矩阵代数的起点。



从对行列式的操作演变成真正的抽象矩阵代数花费了整整46年的时间。尽管行列式的操作及其对称性都是那样地迷人，但是它还是完全依存于数值的世界。一个行列式是一个数，尽管它的计算要求我们仔细观察一个复杂的代数表达式。现代人理解的矩阵不是一个数，它是一个数组，就像我讨论过的数组一样。它的行上和列上的元素可能是数（也不一定必须如此），而且还有一个非常重要的数与它相关，这个数就是它的行列式。然而，对数学家来说矩阵本身就很有意义。总之，它是一个新的数学对象。

我们可以把两个正方形矩阵[7]相加或者相减得到另一个矩阵：只需逐项把它们加起来即可。（虽然两个矩阵相关的行列式出现了不等的情况，但这个运算是适用的。）我们可以做一个数与一个矩阵的乘法得到一个不同的矩阵。这听起来不是很耳熟吗？所有$n \times n$矩阵形成一个向量空间，维度是n^2。另外，可以利用我对行列式所用的相同技术把两个矩阵相乘，因此所有$n \times n$矩阵不仅形成了一个向量空间，而且还形成一个代数！

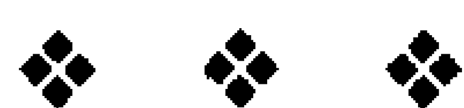
我们可以验证它是所有代数中最重要的一个。例如，它包含很多其他代数。举一个简单例子，复数的代数连带它的加、减、乘和除法法则可以完全映射到2×2矩阵的某个子集上。你可以验证一下，当复数$a+bi$和$c+di$对应于下面两个矩阵时，矩阵的乘法法则（与我在上文所给的行列式乘法法则相同）的确是再生了复数的乘法法则：

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \times \begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} (ac-bd) & (ad+bc) \\ -(ad+bc) & (ac-bd) \end{pmatrix}$$

也许需要确定在这一方案下代表i的矩阵是什么，然后把这个矩阵与其自身相乘，并验证得到的的确是一个代表−1的矩阵。

哈密尔顿四元组可以类似地映射到4×4矩阵。它们的乘法不可交换的事实无关紧要，因为矩阵乘法也不可交换。（尽管有一些特殊的矩阵，例如代表复数的矩阵，它们的乘法具有交换性。）非交换性意想不到地出现在了整个19世纪中期的代数中。置换也是非交换的，它也可以用矩阵表示，而这种数学会让我们的研究更深入。

总之，矩阵是一个很了不起的东西。它非常有用，任何现代代数课程都是从详细介绍矩阵开始的。



我已经说过，行列式演变到矩阵用了46年。1812年柯西在法国科学院宣读了行列式的总结性论文。第一个在代数文献中使用词汇“矩阵”的人是英国数学家西尔维斯特，他在1850年发表的一篇学术论文中使用了这个词。西尔维斯特定义矩阵是“项的长方形排列”。然而，他的思想还是植根于行列式。矩阵本身是一个数学对象的第一次正式认识出现在另一篇论文中，是英国数学家凯莱于1858年发表在《伦敦哲学社会学报》上的，标题为“矩阵理论之论证”。

凯莱和西尔维斯特二人在数学史中总是相提并论，我觉得没有什么理由不遵从这一传统。他们几乎是同时代人，西尔维斯特（出生于1814年）年长7岁。1850年他们相遇，当时二人都在伦敦做实习律师，后来成为亲密的朋友。二人都在研究矩阵，也都在研究不变量（这个出现得更晚些）。二人都在剑桥学习，只是所在的学院不同。

凯莱被选入他所在的学院——剑桥三一学院的院士，并在那里任教四年。由于接下来他不得不在英格兰教堂担任圣职（这显然不是他愿意做的事），所以他转去学习了法律，并于1849年获准成为一名大律师。

西尔维斯特的第一份工作是在伦敦的一所新成立的大学担任自然哲学教授。德·摩根（第10章介绍）是他的同事。但1841年西尔维斯特就离开了，去美国弗吉尼亚大学担任教授。在那里他仅待了三个月，由于与某个学生交恶而辞职。这一事件有多个版本，我不知道哪个是真的。

显然这名学生侮辱了西尔维斯特，但是接下来发生的事情就众说纷纭了。有人说西尔维斯特用剑刺了这名学生，而且拒绝道歉。也有人说西尔维斯特要求学校惩罚这名学生，但是学校没有这样做。甚至还有人说这是同性恋人之间的纠纷。西尔维斯特终生未婚，写华丽的诗，喜欢高音域的演唱，那位在数学上一向攀龙附凤的托马斯·赫斯特在日记里这么说：

> 星期一接到西尔维斯特的信后，我去霍巴特俱乐部看他……他热情得过了头，而且希望我们能够住到一起，请我跟他到乌里芝生活，等等。总之，他对人表现出来的友爱不太正常。

不管真相如何，我觉得我们没有必要去猜测西尔维斯特的私人生活，并由此来搞清楚为什么这个人物的个性、他的犹太人身份（他出生时有另一个名字“约瑟夫”，“西尔维斯特”是后来加上的家族姓）、他的反奴隶制度观点，都无助于他被南北战争前夏洛茨维尔的校园社交圈里的红人们接纳。他后来回到了英格兰，做了保险精算师的工作，学习法律，招收私人学生聊补收入。（其中有一位学生就是弗洛伦斯·南丁格尔，“掌灯女士”，一位很有能力的数学家和统计学家。）

凯莱和西尔维斯特只是19世纪初期英格兰出现的一批优秀代数学家中的两个。当然，哈密尔顿也是这些人中的一员。事实上，有必要为萨克雷笔下某个人物所称的“多雾小岛”单独开设一章，以便更好地了解19世纪初期和中期代数思想的伟大转变的某些背景。

注 解

[1] 这也是古老中国的书写语言有如此严格简洁风格的一个原因。经典教科书就完全不是这样有助于记忆的形式。慎终追远，是孔子告诫我们的话。理各把这句话翻译为：“我们要隆重地举行父母的葬礼仪式，让他们永远享用这供奉的仪式。”（Let there be a careful attention to perform the funeral rites to parents,let them be followed when long gone with the ceremonies of sacrifice.）中文中的四个字，而英语用了39个字节。

[2] 这个年历是汉代天文学家落下闳的作品，他大约生活在公元前130年到公

元前70年。

[3] 乔治·麦克唐纳·罗斯的莱布尼茨传记《莱布尼茨》（牛津大学出版社，1984年）。

[4] 试图得到诸如$1^5+2^5+3^5+\cdots$这样的整数幂求和公式时，就会得到伯努利数。但是准确地解释起来实在太长，我就不介绍了，读者可参看康威和盖伊合著的《数值的书》。从B_0开始的前几个伯努利数是1、−1/2、1/6、0、−1/30、0、1/42、−1/30（是的，又出现了）、0、5/66、0、−691/2730、0、7/6、0、−3617/510、0、436 867/798等。注意，B_1后面第奇数个伯努利数都是零。我将在第12章介绍伯努利数。

[5] 做大量的观察会使得到的结果更精确。另外，行星也不会因行星间的相互作用而越出它们的第二量级理想曲线，所以才有了高斯对小惑星的六次观察。高斯知道克莱姆法则吗？一定知道，但是因为这些都是一些特殊的计算，特殊的排除法就足够了。

[6] 当我还是本科生的时候，老师教我把这（两个矩阵的乘积过程）看成是“从行潜入到列”。为了计算位于积矩阵第m行第n列上的元素，就要取第一个矩阵的第m行，然后把它顺时针方向倾斜90°，再把它沿第二个矩阵的第n列的边下行。把配对的数乘起来，然后再把乘积相加，就得到了想要的元素。例如，下面是一个矩阵积，正确的矩阵记法如下：

$$\begin{pmatrix} 1 & 2 & -1 \\ 3 & 8 & 2 \\ 4 & 9 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & -1 & 4 \\ 7 & 6 & -2 \\ 5 & -1 & -5 \end{pmatrix} = \begin{pmatrix} 10 & 12 & 5 \\ 69 & 43 & -14 \\ 62 & 51 & 3 \end{pmatrix}$$

为了得到答案中第二行、第三列上的−14，我取第一个矩阵的第二行(3，8，2)，然后再取第二个矩阵的第三列(4，−2，−5)，然后把这行“潜入”到这一列计算得$3\times4+8\times(-2)+2\times(-5)=-14$。这就是矩阵的乘法。你可以用其他顺序把这两个矩阵相乘，就可以验证，在一般情况下矩阵乘法是不可交换的。如下所示，把第二个矩阵的第一行潜入到第一个矩阵的第一列，$1\times1+(-1)\times3+4+4=14$，所以这个积矩阵左上角的数是14而不是10。

[7] 矩阵不一定是正方形的。考虑一下矩阵的乘法法则，第一个矩阵的行与第二个矩阵的列结合，就会明白只要第一个矩阵的列数等于第二个矩阵的行数，矩阵的乘法就可行。事实上，一个m行、n列的矩阵可以和一个n行、p列的矩阵相乘，它们的积是一个有m行和p列的矩阵。一个特殊情况就是$p=1$。任意一本比较好的本科生现代代数教材都会澄清这个问题。我还是推荐迈克尔·阿廷的《代数》一书。弗兰克·艾尔斯所著的《矩阵》也非常好。

第10章

维多利亚的多雾小岛

下面是英国数学家皮考克在1830年出版的《代数论》中的一段话：

> [算术]只是一种推测的科学，代数法则和运算也适用于它，但是它们不受算术的限制和制约。

（仿宋体部分是我的强调。）10年后，皮考克的弟子、年轻的苏格兰数学家邓肯·格雷戈里（当时27岁）写了这样一段话：

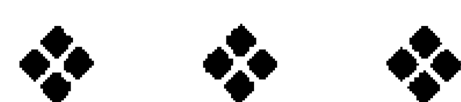
> 在普通代数中有很多定理，虽然表面上看只适用于代表数值的符号，但其实有更广泛的应用。这样的定理只取决于这些符号要服从的结合定律，因此无论它们的性质如何，对于服从结合定律的所有符号这些定理都成立。

下面这段话是另一位英国人德·摩根在他的《三角学与双重代数》（1849年）中所写的：

> 假设符号M、N、+存在唯一一种结合关系，即$(M+N)$与$(N+M)$一样。如何把这种符号计算做成一种有意义的符号计算？有几种方法。(1) M和N可以是数量，+是把第二个量加到第一个量上的加法符号。(2) M和N可以是数值，而+是第二个数乘以第一个数的符号。(3) M和N是直线，+是做以第一项为底、第二项为高的矩形的方向。(4) M和N可以是人，+是前者是后者的兄弟的声

明。(5) M和N可以是国家，+表示后者与前者进行过战斗。

显然在19世纪第二个25年间，代数从数值世界挣脱出来获得了自由。是什么推动了这一过程呢？为什么这些陈述都是出自英伦诸岛的数学家之口呢？



进入18世纪以后，英国的数学越发落后于欧洲大陆数学的发展。艾萨克·牛顿对此要承担部分责任；或者，更恰当的说法是大不列颠民族自我意识膨胀直接导致了这一结果，因为对于他们大部分人来说，牛顿先生是一位民族英雄。这种自我意识的膨胀产生了同样强大的反作用：欧洲大陆各国能够树立与牛顿比肩的文化英雄。笛卡儿就是法国人的英雄。前面提到的帕特丽夏·法拉[1]的著作记录了1760年左右充满爱国思想的大不列颠酒歌，其大意为：

> 牛顿爵士打破了笛卡儿的原子
> 莱布尼茨偷走了我们大不列颠人的微积分
> 牛顿证明了引力，揭示了自然的空间
> 虽然有普莱纽姆结构的偏见
> 我们以万有引力而感到自豪
> 让我举杯祝福
> 比所有这些人都更强大的人——牛顿爵士

事实上，德国有两名反牛顿的代表人物。不仅有莱布尼茨还有歌德，歌德是牛顿光学理论的辛辣的批评家。法拉的书中提到，"歌德在魏玛的房子至今仍然有明显的反牛顿学说的装饰。"[2]

这一切都让大不列颠数学显得很悲哀，因为牛顿设计的微积分运算符号远远不如莱布尼茨设计的符号合理，而且后者为整个欧洲大陆所采用。爱国的大不列颠人坚持牛顿的"点"记号而不接受莱布尼茨的d。例如，英国人写成$\ddot{x}$，而欧陆的写法是

$$\frac{d^2x}{dt^2}$$

这种书写方式使得大不列颠的微分孤立且阻滞不前。[3]对欧洲大陆的数学家来说，这种书写方式令英国人的论文读起来有点烦人，而且掩盖了一个事实，即x（此处是t的函数）受到算子d^2/dt^2的作用：这个算子可以拆开，而且其自身可以看成是一个数学对象。人们已经意识到了这个事实的重要性。

即使承认是牛顿的原因，也不可避免给人一种这样的感觉：顽固的民族自尊心和狭隘的岛国意识正在拖英国数学的后腿。例如，复数已经"定居"欧洲数学相当长的时间了，而在英国甚至负数也要受到某些职业数学家们的嘲笑，下面这段话就是一个佐证，它取自威廉•弗雷德的《代数原理》（1796年）一书。

> [一个数]可以从一个比它大的数减去它，但努力尝试从一个比它小的数减去它是可笑的。然而，那些谈论负数，谈论把一个负数与一个负数相乘因而得到一个正数，谈论虚构的数的代数学家们正在做这种努力。这完全是胡言乱语，是违背常识的。但是一旦它被接受，就如其他很多虚构的事情一样，它很快就会在那些以某些事物为信仰并憎恨严肃思想的人中间得到巨大的支持。

弗雷德不是一个孤立的行为古怪之人。他曾经是1780年剑桥大学基督学院数学考试的第二名。他是英国第一批精算师，后来与我在前文介绍过的德•摩根成为了朋友，而且还把自己的一个女儿嫁给了德•摩根。

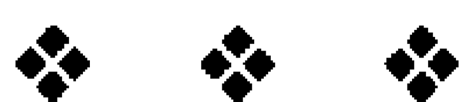
到了19世纪初期，英国年轻一代数学家已经开始对事态感到不满。抵抗拿破仑的长年战争对英国人产生了双重影响，一方面迫使英国人比以往更加关注欧洲大陆思想，另一方面也使得这个岛国的数学家认识到法国的数学是多么优秀。

1813年，牛顿曾经就学的三一学院的三个年轻学者采取了具体行动，建立了他们所谓的分析社团。这三个学者都出生于1791年或1792年，分别是发现天王星的天文学家之子约翰·赫歇耳，后来因为其“计算工具”（一种简单的机械计算机）而出名的查尔斯·巴贝奇，以及我在前面的引文的作者乔治·皮考克。他们社团的主要宗旨是改革微积分教学，正如巴贝奇一语双关说的那样，“以纯d主义的原理对抗大学的点（dot）时代。”

然而，分析社团似乎没有持续很长时间，他们三人都没有成为一流的数学家，但是这个社团的精神被皮考克这位积极的唯心主义改革者以他那个时代特有的风格发扬光大。毕业后，他在三一学院担任讲师，并于1817年成为一名数学考官。他被任命为考试官后的第一件事就是改革微积分教学，把教学从牛顿的“点”时代变成莱布尼茨的d主义。

皮考克后来成为一名全职教授，为几个学术团体的创建发挥了作用，特别是为伦敦天文社、剑桥哲学社和英国科学促进协会的创建做出了贡献。所有这些新团体都面向有能力和有成就的任何人，打破了原来认为学术团体只是绅士俱乐部、拒绝接受那些自学成才的劳动人民和“粗鲁工匠”等的老观念。工业革命初期，持技术专家治国论的低中阶层展示了他们的力量。皮考克仕途的最后一站是英格兰东部的伊利教堂主持，他对自己的一生颇为满意。

这种整体的改革精神是这一时期英国数学变革的背景。其成果可以在下一代英国数学家，特别是代数学家身上看到。我已经介绍了出生于1805年的哈密尔顿的工作。紧接着就是德·摩根（出生于1806年）、西尔维斯特（出生于1814年）、布尔（出生于1815年）和凯莱（出生于1821年）。这些人重新为英国的数学赢得了荣誉，至少在代数学上是这样。我们要把群论、矩阵理论、不变量理论以及数学基础的现代理论的全部或一部分归功于他们。[4]



德·摩根是我刚才提到的四位数学家当中数学成果最少的人，但是在很多方面是最有趣的一个人。他在我心中有着特殊的位置，因为他是我的母校伦敦大学学院的第一位数学教授，这所大学是“高尔街上的无神论学院”。

像牛津和剑桥这样历史悠久的英国名校，进入19世纪以来仍然受到其早期历史遗留下来的各种宗教、社会和政治等方方面面的习俗的制约。例如，他们都不接收女学生，都要求对硕士和会士有一次宗教考试（牛津实际上要求对本科生也进行这样的考试），主要是要他们宣誓对英国教堂及其教义忠诚。这些限制[5]对大数人来说渐渐变得很可笑，因此拿破仑之战伊始，改革意识开始流行时，英国的知识阶层也马上掀起了促进英国高等教育进步的热浪。这一热浪在伦敦大学的建立过程中得到了充分的体现，1828年这所学校开始招收第一批学生。

这所新大学[6]是英格兰第一所不限制性别、不限制宗教信仰和政治观点的高等学府。很快就有更多这样的学院在这座城市建立起来了。在21世纪初，这所大学如其他古老大学一样是一所综合性大学，而最初在高尔街上建立的学院就是现在的伦敦大学学院。

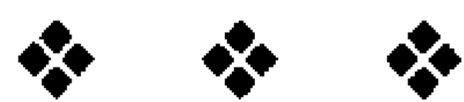
这所新大学的建立对德·摩根来说正是时候。他在三一学院取得了学士学位。皮考克曾是他的老师，另外还有乔治·艾里，此人后来成为皇家天文学家，有以他的名字命名的数学函数[7]。在1826年进行的毕业数学考试取得第四名之后，德·摩根打算攻读硕士学位。然而，当时要求宗教考试。德·摩根似乎有一种天生的（他的传记作家杰文斯说“深深的”）宗教倾向，但是他的信仰是个人的，没有参加任何有组织的教堂，当然也没有参加英国大教堂。

德·摩根一直是一个原则性很强的人，他拒绝参加这样的考试，回到了伦敦的家，就像20年后的凯莱一样，自动退学想成为一名律师。他刚刚在林肯律师学院登记注册，这所新大学就成立了，并为他提供了数

学教授的职位。他接受了。他22岁时，在高尔街开始了他的第一堂课，“论数学学习”。此后德·摩根一直在这个岗位上尽职，直到1866年，他因一个原则性问题辞职。

德·摩根好学、和蔼，深深打动了阅读他的传记的读者们，他就是那种人们愿意邀请共进晚宴的人。他还是一位伟大的科学普及者，热情地向很多社团和小型杂志投稿，满足后乔治时代和维多利亚初期不断壮大的技术和商业阶层人士的需求。（“他写的文章不少于850篇。”他的传记作家说。）他是一位藏书家和优秀的业余长笛手。他的妻子在他家（切尔西的切尼步道30号）附近经营了一家法国古典风格的知识分子沙龙。他的女儿是童话作家。我至今还记得她写的那个令人毛骨悚然的童话“头发树”，那是我孩童时代最害怕的东西。

德·摩根善于解决难题，对语言和数学方面的怪异事情有着强烈的热忱，其中的一些事件收集在他的佳作《一组悖论》中（在他1872年去世后由他妻子出版）。当他知道他在x^2那一年他正好x岁（只有极少数人能幸运地赶上这个巧合[8]），并且自己的名字（Augustus De Morgan）是“O Gus! Tug a mean surd!”①的相同字母异序词时特别高兴。



德·摩根对代数历史的重要贡献在于他尝试完善逻辑并改进它的记法。自从亚里士多德开创逻辑以来，它几乎没有任何发展。到了德·摩根时代开始教授逻辑，它依然以这样的思想为基础：一共有四种类型的命题，两种肯定命题，两种否定命题。这些类型是：

完全肯定（所有X都是Y）

部分肯定（某些X是Y）

完全否定（X不是Y）

部分否定（某些X不是Y）

---

①噢，格斯！抓住一个均值根式。（格斯是Augustus的昵称。）——编者注

这样的命题可以三个组成一组，称为三段论，两个命题可以得出一个结论：

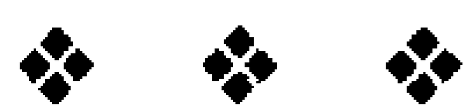
所有人都会死

苏格拉底是人

苏格拉底会死

在德·摩根的时代，爱丁堡有一位逻辑和形而上学教授威廉·哈密尔顿，这与第8章中介绍的威廉·哈密尔顿不是同一个人，然而这两个人总被弄混。在1833年的一次逻辑课上，这位“另外一个哈密尔顿”建议改进亚里士多德的系统。他认为亚里士多德量化他的命题主语（所有的X、某些X）而不量化它的谓语（是Y、不是Y）是错误的。他建议谓语量化。

德·摩根接受了这一观点，并采用了它，最终出版了一本名为《形式逻辑，或者推论的演算，必要性和可能性》的书（1847年）。在这本书出版后的几年间，他又相继发表了四个论文集，并打算把这些著作综合起来建立一个良好的记法和谓词量化的巨大新逻辑体系。提出原始思想的威廉·哈密尔顿没有受到什么触动。他认为德·摩根的体系“带有很多可怕的小刺”。今天看来这只是一件历史趣闻，因为德·摩根只是改进了书写逻辑公式的传统方法。逻辑进步真正需要的是完整的现代代数符号体系，这是由布尔完成的。



布尔是19世纪初英国的“新人类”。他出身卑微，自学成才，没有得到什么人的资助，什么也没有，只靠他自己的能力和努力提升自己。他是小镇皮匠和女仆的儿子，他的父母竭尽所能也只能给他提供有限的正规教育，他只能靠自己刻苦钻研来弥补不足。14岁时，他开始翻译希腊诗。16岁时，父亲破产，他不得不寻找一份教师的工作来支撑这个家。他做了18年教师工作，大部分时间是经营他自己的学校。早在19岁那年，他就创办了第一所自己的学校。

他大约在17岁开始认真学习数学，并且很快就掌握了微积分。二十几岁的时候，在《剑桥数学杂志》第一编辑邓肯·格雷戈里的鼓励下，他开始定期在这本杂志上发表文章。我在本章开始时引用了这个人的话。1842年他开始与德·摩根通信，而德·摩根帮助布尔在皇家科学院发表了一篇关于微分方程的论文。

1846年，当英国政府宣布在爱尔兰扩大高等教育的时候，德·摩根、凯莱和威廉·托马斯（后来的开尔文勋爵，温标就是以他的名字命名的）这些欣赏布尔的人为布尔极力争取其中一所新学院的教授职位。他们成功了，1849年布尔成为科克城女王学院的数学教授，并且一干就是15年。直到1864年11月的一天，他冒着大雨步行两英里从家里走到学校，穿着湿衣服上课，结果感冒了。他的妻子认为着凉感冒也要“以毒攻毒”，因此她让布尔躺在床上，然后用一桶桶冷水浇他。正如数学家们常说的那样，结果很显然。

我没有找到任何对布尔不友好的言论。即使对赞扬他的传记作家的过分夸张打个折扣，他似乎还是一位近乎圣洁的好人。他娶了乔治·埃佛勒斯［珠穆朗玛峰（Everest）就是以这个人的名字命名的］爵士的侄女，生活幸福。夫妇两育有5个女儿。他们的第三个女儿艾丽西娅·布尔·斯托德[9]自学数学，在多维几何方面做出了重要贡献。她一直活到了80岁，在第二次世界大战时死于英格兰。

布尔的伟大成就是逻辑的代数化，通过使用代数符号提高了逻辑的地位，使其成为数学的一个分支。为了说明布尔的方法，下面给出我已经介绍过的三段论法的代数版本。

假设生活在地球上的所有生物是一个集合。这就是我们的“论域”，只是布尔没有使用这样的术语，一直到1881年才由维恩发明了这一术语。设这个论域是1。在同样意义下，用0表示空集，这个集合根本没有任何成员。现在考虑所有会死的生物，用x表示这个集合。（有可能$x=1$，这不影响讨论。）类似地，用y表示所有人的集合，s是只有一个成员（即苏格

拉底）的集合。

另外还有两种记法。第一，如果p是事物的集合，且q也是事物的集合，那么乘法符号就表示他们的交集——$p \times q$代表既在p中又在q中的那些事物。（也可能没有这样的事物存在，即$p \times q=0$。）第二，利用减法符号表示去除这个交集：$p-q$表示所有在p中但不在q中的事物。

现在，将这三段论代数化。“所有人都会死”这句话可以重新表述为“是人且不死的生物集合是一个空集”。使用代数表示就是：$y \times (1-x)=0$。乘开括号并运用花拉子米的法则，这个表达式等于$y=y \times x$。（翻译出来就是“所有人的集合正好等于会死的人的集合”。）

“苏格拉底是一个人”类似地代数化为$s \times (1-y)=0$，相当于$s=s \times y$。（“刚好由苏格拉底组成的集合等于其成员只有一个且同时是苏格拉底和人的集合。”如果苏格拉底不是人，上面的等式就不成立，后面的集合就是空的！）

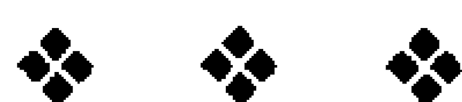
将$y=y \times x$代入等式$s=s \times y$，得到$s=s \times (y \times x)$。根据一般的代数法则，我可以这样重新定位括号的位置：$s=(s \times y) \times x$。但是我已经证明$s \times y$等于s，因此$s=s \times x$等价于$s \times (1-x)=0$。翻译过来就是：“是苏格拉底且不会死的生物的集合是空集。”所以苏格拉底会死。

伯特兰·罗素在1901年发表的一篇文章中有一段被广为引用的评论：“纯数学是由布尔发明的，发表在名为《思维规律》的一本著作中。”罗素进而在那一评论中表明了自己的态度：“如果他的著作真包含了思维规律，那么很奇怪，之前没有一个人用这种方式思维……”罗素还根据他自己对数学和逻辑之间的关系的认识做出结论：数学就是逻辑。这是一个今天已不被广泛认可的信念。

大多数现代数学家对罗素评论的反映应该是，认为布尔真正发明的不是纯数学，而是应用数学的一个新分支，是代数到逻辑的应用。在布尔思想出现之后，历史证明了这一点。通过后面几代逻辑学家的努力，19世纪末，布尔的集合代数已经完全转化成逻辑演算。这些逻辑学家是

休·麦科尔、查尔斯·桑德斯·皮尔斯（他是第8章提到的本杰明的儿子）、朱塞佩·皮亚诺和哥特洛布·弗雷格。于是这种逻辑演算成为20世纪探究的潮流，在数学系的课程表上均属于基础课，在此我们使用数学技术来研究数学本身的性质。

因为通常不把这一潮流看成是现代数学的一部分，所以我不再进一步陈述了。然而，代数的历史会因为没有布尔代数的某些记述而变得不完整，乔治·布尔功勋卓著，是他把代数与逻辑结合在了一起。



作为出生于19世纪第一个25年的英国一代伟大数学家，凯莱已经在我介绍矩阵的相关理论时小露了一次脸。然而这并非凯莱对代数做出的唯一贡献。他有权声明自己是现代抽象群论（下一章介绍）的创始人。因此在介绍欧洲大陆数学发展之前，正好在这里陈述一下凯莱的工作，这对他来说也是公平的。

英语单词“群”在现代代数的意义上的第一次出现是在凯莱于1854年发表的两篇论文中，这两篇论文的标题相同——“条件 $\theta^n=1$ 下的群论”。下一章将详细介绍群论的初期发展，在这里我只想给出凯莱在1854年的描述中的一个群论的有用特征。

回想第9章讨论行列式时，我特别研究了3个对象的置换，列出了这样置换的所有6种可能情况。为了解释凯莱的发展，我需要更详细地讲述置换的相关内容并介绍一种书写它们的好方法。曾经有三四种表示置换的不同方法，但是现代代数学家似乎完全倾向于**循环记法**，这就是从现在开始我要使用的记法。

循环记法是这样的。考虑标有1、2和3的3个盒子里的3种对象：苹果、书和梳子。把下面的状态看成是一个“起始状态”：苹果在盒子1里，书在盒子2里，梳子在盒子3里。定义“恒等置换”为不做任何改变的置换。如果把恒等置换运用于起始状态，那么苹果仍然在盒子1里，书仍在盒子2里，梳子仍在盒子3里。

注意，下面这一点经常令初学者感到困惑：无论盒子里当时是什么东西，置换的作用对象都是盒子里的内容。循环记法(12)表示如下置换：交换第一个盒子里的内容和第二个盒子里的内容。其意义是：[盒子]1[的对象]进入[盒子]2，[盒子]2[的对象]进入[盒子]1。正如方括号所显示的那样，当一位数学家看到这个循环记法时实际想的是：1走到2，2走到1。注意括号的效果，括号中的最后的数字置换到第一个位置。这就是为什么这个记法叫做循环记法的原因！

假设我们把这一置换作用于初始状态。那么，苹果将在盒子2里，而书将在盒子1里。如果再用不做任何事情的恒等变换，苹果仍在盒子2里，书仍在盒子1里，梳子仍在盒子3里。使用乘法符号表示置换的合成，(12)×I=(12)。

假设我从初始状态做(12)置换，然后再做一次。第一次置换后，苹果到了盒子2里，书到了盒子1里。第二次置换后，书到了盒子2里，苹果到了盒子1里。我又回到了初始状态。换句话说，(12)×(12)=I。

如果像上面这样处理，就可以构建一个完整的三元素置换“乘法表”。用循环记法(132)来表示更复杂的置换，其意义是：[盒子]1[的对象]进入[盒子]3，[盒子]3[的对象]进入[盒子]2，[盒子]2[的对象]进入[盒子]1。好，对初始状态做这一置换，结果是苹果进入盒子3里，梳子进入盒子2里，而书则进入盒子1里（1到3，3到2，2到1）。然后我再做置换(12)，于是梳子进入盒子1里，书进入盒子2里，苹果进入盒子3里，结果跟对初始状态做置换(13)一样。用代数表示：(132)×(12)=(13)。

我说过，可以像这样建立一个完整的乘法表，如图10-1所示。先做左边列表的置换，再做表上方列表中的置换，那么作用结果就在前一个置换所在行和后一个置换所在列的交叉位置。

这个表称为凯莱表。事实上，这个表与1854年的那两篇文章中的第一篇文章的第6页上的一个表非常相近。注意置换的合成是不可交换的，这一点可以从这个表上相对主对角线(从左上角到右下角)不对称看出来。

	I	(123)	(132)	(23)	(13)	(12)
I	I	(123)	(132)	(23)	(13)	(12)
(123)	(123)	(132)	I	(13)	(12)	(23)
(132)	(132)	I	(123)	(12)	(23)	(13)
(23)	(23)	(12)	(13)	I	(132)	(123)
(13)	(13)	(23)	(12)	(123)	I	(132)
(12)	(12)	(13)	(23)	(132)	(123)	I

图10-1 群S_3的凯莱表

3个对象的这6种可能置换，再加上图10-1的表所定义的合成置换的法则，就是一个群的例子。这个群非常重要，因此它有自己的符号S_3。

S_3不是唯一的六元素群，还有另外一个群。考虑单位的六次方根。仍用ω表示第一个立方根，则单位的六次方根是：1，$-\omega^2$，ω，-1，ω^2，$-\omega$。（参见“数学知识：单位根”。）这6个数的一般乘法表见图10-2。

	1	$-\omega^2$	ω	-1	ω^2	$-\omega$
1	1	$-\omega^2$	ω	-1	ω^2	$-\omega$
$-\omega^2$	$-\omega^2$	ω	-1	ω^2	$-\omega$	1
ω	ω	-1	ω^2	$-\omega$	1	$-\omega^2$
-1	-1	ω^2	$-\omega$	1	$-\omega^2$	ω
ω^2	ω^2	$-\omega$	1	$-\omega^2$	ω	-1
$-\omega$	$-\omega$	1	$-\omega^2$	ω	-1	ω^2

图10-2 群C_6的凯莱表

跟预想的一样，这个群是可交换的，因为我只是在做普通的（我的意思就是普通复数的）乘法。它的名字是C_6，是一个六元素循环群。

这些都是六元素群的例子，用术语表示，它们都是秩为6的群。为什么说它们是群呢？乘法表的某些特征是关键点。例如，注意单位元（第一个群的单位元是I，第二个群的单位元是1）在每一行和每一列刚好出

现一次。

上一段话中最重要的一个词就是“例子”。数学家会认为S_3和C_6这两个秩为6的群是完全抽象的对象。如果把3个对象的6个置换替换成为1、α、β、γ、δ和ε，并且用相应的希腊字母替换凯莱表中的每一个符号，那么这个表格就表示一个抽象群的定义，它与置换完全无关了。事实上，这就是凯莱在1854年的两篇论文中的做法。3个对象的置换群是抽象群S_3的一个实例，就像美国最高法院的法官人数是抽象数字9的一个实例一样。单位的六次方根的情况也类似，对它们做普通乘法运算，就构成了C_6。可以用1、α、β等做替换，并对第二个表格也做相应的替换，就得到C_6的完全抽象的定义，与单位根没有任何关系。

这就是凯莱的伟大成就，用这样纯抽象的方式展示群的思想。尽管凯莱有如此的真知灼见，而且我们也承认他于1854年发表的两篇论文所表现出来的概念上的巨大飞跃，但是凯莱还是不能完全把他的研究对象从方程及其根的研究背景中摆脱出来。凯莱回顾了这些背景，并在第一篇论文的第二页附加了这样的脚注：

> 是伽罗华提出了应用于置换或替换的群的概念，可以认为群的引入开创了代数方程理论发展的新纪元。

凯莱是完全正确的。现在我们该回过头重新拾起19世纪中间25年的另一条线索，去看一看代数史上唯一具有浪漫色彩的英雄人物伽罗华。

注　解

[1] 参看第6章注释[2]。

[2] 德·摩根在他的著作《奇人的预算》中评论类似的主题的不同酒歌时说：“1800年，赞美之词只给了牛顿而没有给笛卡儿也许是一种失衡的结构。”

[3] 英国人珍爱牛顿的微积分记法，使其苟延残喘，至少在教科书中是这样的。20世纪60年代初在一所优秀的英国男校，我学习了利用牛顿点记法的物理和应用数学。

[4] 我引用过引语的这位苏格兰人邓肯·乔治全身心地投入这一时期的数学评论。他在布尔出生三年多后就去世了，但他对布尔却产生了很大的影响。事实上，我提到的邓肯的引语不是取自于《爱丁堡皇家学会学报》的原版，而是取自于一篇论文（“论分析中的一般方法”），这篇论文是布尔于1844年向皇家学会提交的。

[5] 当然还有税收。在对抗拿破仑的战争中为备不时之需，开始执行税收政策。战争结束时，改革者亨利·伯洛格哈姆劝说国会税收没有必要再持续。这让政府恐慌，但是却使人民普遍感到高兴。1816年这一税收被废除了。

[6] “伦敦大学是犹太人和威尔士人创建的。”当我第一次走入这所学校的大门时，人们就是这样告诉我的。事实上，詹姆斯·穆勒、托马斯·坎贝尔、亨利·伯洛格哈姆都是苏格兰人，他们是这所大学建立的幕后精神支持者。然而，为建立这所大学筹集而来的资金都来自于这个城市的商人阶层，其中大部分人的确是卫理公会派教徒和犹太人。

[7] 这是两个关系很密切的函数。它们是下面常微分方程的解，而且在很多物理学分支中出现：

$$d^2y/d^2x=xy$$

[8] 只有出生在（N^2-N）这一年才能这么幸运。德·摩根出生于1806年（$N=43$）。其后的幸运出生年份是1892年、1980年和2070年。

[9] 在20世纪30年代，当时已经70多岁高龄的艾丽西娅与伟大的几何学家考克斯特（1907—2002）一起工作。考克斯特在他的《正规多边形》中对她的介绍颇多：“当她四岁的时候，她的父亲就去世了，所以她的数学才能纯粹是遗传……她接受的不可能是普通的教育，布尔先生与（神秘主义者、物理学家、古怪之人、社会激进分子）詹姆斯·辛顿是朋友，这一关系把社会上的改革者和奇怪之人都吸引到了这个家庭。那几年间，辛顿的儿子霍华德带来了很多木制的小立方体，他让布尔的这个最小的三女儿记住他给这些小木块随意起的拉丁语名字，或者叫她把这些小木块堆成各种形状。这给当时大约8岁的艾丽西娅带来了创建四维几何的非凡灵感……”顺便提一下，那个霍华德的全名是查尔斯·霍华德·辛顿，关于第四维的某些推断就是他提出来的，这也许为艾勃特创作《二维国》带来了灵感。

第三部分

抽 象 层 次

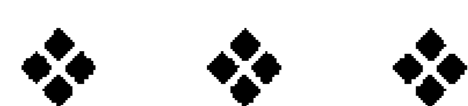
数学知识

域　论

域和群是19世纪初期一系列研究中发现的两个数学对象的名字。

从内部结构来看，域比群更复杂。因此，代数课本通常先介绍群，然后再提升到域。但我们感觉域比群更普通，更容易理解。群因为更简单，有着更广泛的适用性，所以大体上，对纯代数学家来说，群论比域论更具挑战性[1]。

因为这些原因，也为了使伽罗华的发现更容易理解，在更详细介绍群论之前，我在这里介绍一下域。



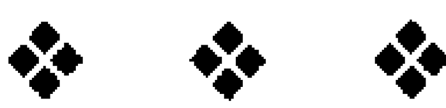
域是一个可以随意进行加、减、乘和除的数系（或者其他东西，但目前讲的是数）。无论做多少次这四种基本运算，答案总是同一个域里的数。这就是为什么我说域很普通。在域中，所做的就是基本算术：+，−，×，÷。如果想得到域这一代数概念的一个形象的记忆方法，那么就想象一下有加、减、乘、除这四个操作键的最简易的计算器。

在域中需要遵守几个法则，这些法则都很常见。我已经介绍了封闭性法则：算术运算的结果仍在这个域里。还需要一个“零”，这样才能把它与其他数相加时保持这个数不变。另外还需要一个1，当其他数与它相乘时保持不变。必须遵守基本的代数法则：$a\times(b+c)$总是等于$a\times b+a\times c$。加法和乘法必须是可交换的，在域中我们不与不可交换性打交道。因此，哈密尔顿的四元数不是一个域，只是一个“除代数”。

$\mathbb{N}$ 和 $\mathbb{Z}$ 都不是域，因为两个整数相除不能总是得出一个整数。然而，

有理数家族$\mathbb{Q}$是一个域。你可以做任意次加、减、乘和除，结果总是在$\mathbb{Q}$中，所以它是一个域。可以说$\mathbb{Q}$是最重要、最基本的域。实数$\mathbb{R}$也形成一个域。运用我在本书一开始给出的加、减、乘和除的复数运算法则，可以验证复数$\mathbb{C}$也是一个域。因此，我们已经知道了三个域的例子：$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$。

除了$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$之外，还有其他域吗？当然有。我将给出两个常见的类型，然后把它们放到一起引出伽罗华理论和群论。最后我会补充介绍第三种重要的域。



除了熟悉的$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$之外，首先介绍的另一种类型的域是有限域。$\mathbb{Q}$、$\mathbb{R}$和$\mathbb{C}$都有无穷多个元素：无穷多个有理数，无穷多个实数，无穷多个复数。

下面是一个只有三个元素的域，为了方便起见，我称它们为0、1和2，当然如果你觉得这些名字容易与更普通的数字混淆，你可以用其他符号替换我的0、1和2。比如，用Z代表0，I代表1，T代表这个域中的第三个元素。例如，在这个域中不会有2+2=4的情况发生。在这个域中，2+2=1。图FT-1所示为这个有限域例子的加法和乘法表，这个域的名字是F_3。

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

×	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

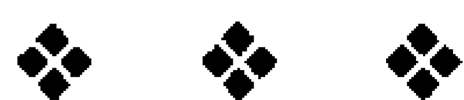
图FT-1 域F_3

注意这个域中的某些要点。首先，因为1的加法的逆元素（即减法）是2，反之亦然，所以对于减法没有什么可说的。减1总可以由加2代替，反之亦然。[2]对于除法也一样，因为2的乘法的逆元素是2（即2×2=1），所以除以2的除法总可以用乘以2的乘法表示，结果相同！除以1的除法就

更简单了，域里不允许有除以零的除法。

对于每一个大于1的自然数，存在相应个数的有限域吗？不。只有素数和它们的幂才有有限域。存在有2、4、8、16、32等个元素的有限域，也存在有3、9、27、81、243等个元素的有限域，依此类推。然而，不存在有6个或者15个元素的有限域。

有限域通常称为伽罗华域，以纪念法国数学家伽罗华，我们在正文中再介绍他。



第二种其他类型的域是扩张域。取一个我们熟悉的域，最通常的就是 $\mathbb{Q}$，向其中再加入一个元素。当然，这个后加入的元素取自这个域之外。

例如，假设我们向 $\mathbb{Q}$ 中加入 $\sqrt{2}$。因为 $\sqrt{2}$ 不在 $\mathbb{Q}$ 中，所以它就是我所说的“后加入的元素”。如果现在我在这个扩大了的数的家族中做加、减、乘和除，都会得到形如（$a+b\sqrt{2}$）的数，其中a和b是有理数。这种形式的两个数的和、差、积和商仍然是这种类型的数。事实上，这些数的加、减、乘和除法的法则有点像复数的相应法则。例如，下面是除法法则：

$$\left(a+b\sqrt{2}\right)\div\left(c+d\sqrt{2}\right)=\frac{ac-2bd}{c^2-2d^2}+\frac{bc-ad}{c^2-2d^2}\sqrt{2}$$

这是一个域。我通过向有理数域 $\mathbb{Q}$ 加入一个无理数 $\sqrt{2}$ 得到了一个扩张了的域，一个新域。

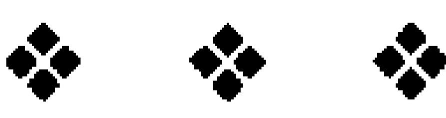
注意，这个新域不是实数域 $\mathbb{R}$。下面这样的实数就不在其中：$\sqrt{3}$、$\sqrt[5]{12}$、π 以及其他无穷多个数都不在其中。在其中的数是这样的一些数：所有有理数，$\sqrt{2}$，所有通过四个基本运算将有理数与 $\sqrt{2}$ 进行合成而得到的数。

我为什么要花费这么大的精力把 $\mathbb{Q}$ 扩张这么一点点呢？其原因就是为了解方程。毕达哥拉斯发现方程$x^2-2=0$在 $\mathbb{Q}$ 中没有解时，他很吃惊，也很沮丧。但是，用这种方法把域稍微扩张一下，它就有了解$x=\sqrt{2}$ 及 $x=-\sqrt{2}$。通过适当地扩张域，我就能够求解之前不能求解的方程。

注意，有一个有趣的要点：上面这个扩张了的域是原始域$\mathbb{Q}$上的向量空间。1和$\sqrt{2}$就是其中两个线性无关的数。它们是这个向量空间的一对很好的基（参见对向量空间的介绍），其他每一个向量，即形如$a+b\sqrt{2}$的数（其中a和b是有理数）都可以用它们来表示。因此，作为向量空间，这个扩张了的域是二维的。

通过向$\mathbb{Q}$加入一个无理数而得到的域并不总是二维的。例如，如果加入的无理数是$\sqrt[3]{2}$，那么这个扩张域是三维的，其中1、$\sqrt[3]{2}$和$\sqrt[3]{4}$是一个基。为了展示事物失控的速度有多快，看看下面这个域的除法法则：

$$\left(a+b\sqrt[3]{2}+c\sqrt[3]{4}\right)\div\left(f+g\sqrt[3]{2}+h\sqrt[3]{4}\right)$$
$$=\frac{af^2-2agh+2bg^2-2bfh+4ch^2-2cfg}{f^3+2g^3+4h^3-6fgh}$$
$$+\frac{2ah^2-afg+bf^2-2bgh+2cg^2-2cfh}{f^3+2g^3+4h^3-6fgh}\sqrt[3]{2}$$
$$+\frac{ag^2-afh+2bh^2-bfg+cf^2-2cgh}{f^3+2g^3+4h^3-6fgh}\sqrt[3]{4}$$



现在我要把前两节的内容结合起来，并在我的0、1、2域求解三次方程。你会看到，有限域的优势是可以写出所有可能的三次方程！

重要的事先做。首先给出我的0、1、2域上的所有可能的线性方程以及它们的解，对于这些解你可以对照前面的加法和乘法表加以核查。

方程	解
$x=0$	$x=0$
$2x=0$	$x=0$
$x+1=0$	$x=2$
$x+2=0$	$x=1$
$2x+1=0$	$x=1$
$2x+2=0$	$x=2$

事实上，我甚至已经列多了。前面两个方程太直白了。显然$2x=0$的

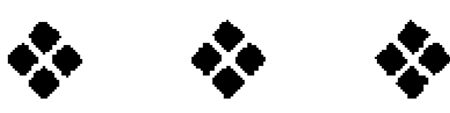
解是x=0！最后两个方程也比较显然。其左边的因式分解分别是2(x+2)和2(x+1)，所以实际上它们分别就是稍加变形的第三个和第四个方程。因此只有中间两个方程有一定的意义。

下面是二次方程。这一次我一开始就不考虑那些没有实际意义的方程。下面是系数属于F_3的所有有实际意义的二次方程。另外，我已经对它们进行了因式分解。

方程	因式分解	解
$x^2+1=0$	不可分解	无解
$x^2+2=0$	$(x+1)(x+2)$	$x=1, x=2$
$x^2+x+1=0$	$(x+2)^2$	$x=1$
$x^2+x+2=0$	不可分解	无解
$x^2+2x+1=0$	$(x+1)^2$	$x=2$
$x^2+2x+2=0$	不可分解	无解

在我研究的这个域内没有解的方程称为不可约的。（比较“数学知识：三次方程和四次方程”中的注解[3]。）你可以看到系数在0、1、2域中的六个有意义的方程中有三个是不可约的。

看一下我都做了什么。我已经在小范围内再建了可以“正常”考虑的二次方程的情况，不同的是，在这里不用考虑无限多个二次方程，这个域上只有六个二次方程：三个有解，三个不可约。在普通算术中，方程$x^2-2=0$在$\mathbb{Q}$内没有解，因为$\sqrt{2}$不是有理数。类似地，方程$x^2+1=0$在$\mathbb{Q}$中没有解，甚至在$\mathbb{R}$中也没有解，因为$\sqrt{-1}$既不在$\mathbb{Q}$内也不在$\mathbb{R}$内。



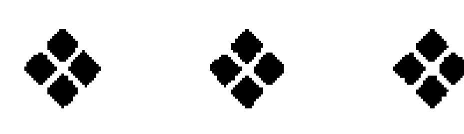
我们能否扩张0、1、2的域，使得那些不可约方程有解呢？是的，这样做是行得通的。发明一个新数，我称其为a，它满足第一个方程：$a^2+1=0$。在这个方程两边加2，得$a^2=2$。（所以，你可以称a是2的一个平方根。因为这个2与通常的2的性质不同，所以我没把a写成$\sqrt{2}$，仍然写成a。）现在，所有的方程都可以求解：

方程	因式分解	解
$x^2+1=0$	$(x+2a)(x+a)$	$x=a, x=2a$
$x^2+2=0$	$(x+1)(x+2)$	$x=1, x=2$
$x^2+x+1=0$	$(x+2)^2$	$x=1$
$x^2+x+2=0$	$(x+2a+2)(x+a+2)$	$x=a+1, x=2a+1$
$x^2+2x+1=0$	$(x+1)^2$	$x=2$
$x^2+2x+2=0$	$(x+2a+1)(x+a+1)$	$x=a+2, x=2a+2$

我们只需要向这个域加入一个元素a，就可以求解所有方程。有限域内的加法、减法、乘法和除法［通常记为$F_3(a)$］只由a的线性表达式构成。如果一个乘法的结果是a^2，就可以立即用2替换它，因为a^2=2。图FT-2所示是这个扩张域的乘法表。（加法表没有什么意思，不过如果你想要一个，可以尝试着自己创建。）

×	0	1	2	a	$2a$	$1+a$	$1+2a$	$2+a$	$2+2a$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	a	$2a$	$1+a$	$1+2a$	$2+a$	$2+2a$
2	0	2	1	$2a$	a	$2+2a$	$2+a$	$1+2a$	$1+a$
a	0	a	$2a$	2	1	$2+a$	$1+a$	$2+2a$	$1+2a$
$2a$	0	$2a$	a	1	2	$1+2a$	$2+2a$	$1+a$	$2+a$
$1+a$	0	$1+a$	$2+2a$	$2+a$	$1+2a$	$2a$	2	1	a
$1+2a$	0	$1+2a$	$2+a$	$1+a$	$2+2a$	2	a	$2a$	1
$2+a$	0	$2+a$	$2+2a$	$2+2a$	$1+a$	1	$2a$	a	2
$2+2a$	0	$2+2a$	$1+a$	$1+2a$	$2+a$	a	1	2	$2a$

图FT-2　$F_3(a)$的乘法表



以上就是伽罗华理论的精髓部分。有一个方程，它的系数属于某个

域，但是它在这个域里没有解。为了囊括这些解，把这个系数域扩大到一个更大的域，称其为解域。伽罗华所关注的是，我们的方程所取的解的形式取决于系数域和解域这两个域之间的关系。

这就是伽罗华的高明之处。1830年他发现这一关系可以用群论的语言（置换语言）表示。

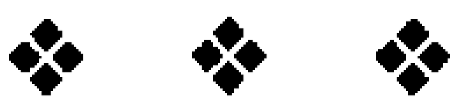
伽罗华发现，对于任意给定的方程，我们需要考虑这个解域的某些置换。像上面我的$F_3(a)$这样的解域一般都比系数域大（在我的例子中系数域是F_3）。现在，在解域诸多置换当中，存在某个置换的子系列，它保持系数域不变。该子系列形成一个群，我们称它是该方程的伽罗华群。所有关于方程可解性的问题都可以转化为关于这个群的结构问题。

对于本节开始我所使用的系数取自于域F_3的方程$x^2+1=0$来说，它的伽罗华群是一个相当简单的群，只有两个成员。其中之一是恒等置换I，也就是什么都不变。另一个置换交换两个解，把a送到$2a$，把$2a$送到a。我们称这个置换为P，它作用于整个F_3(a)。使用箭头表示“置换到”时，这一置换的作用如下所示：$0\to0$，$1\to1$，$2\to2$，$a\to2a$，$2a\to a$，$1+a\to1+2a$，$1+2a\to1+a$，$2+a\to2+2a$，$2+2a\to2+a$。

图FT-3所示是系数域F_3上方程$x^2+1=0$的伽罗华群的乘法表。这里的乘法是置换的合成，即先做一个置换，再做另一个置换。

	I	P
I	I	P
P	P	I

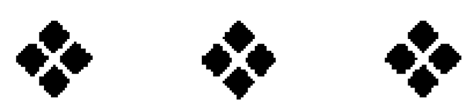
图FT-3　方程$x^2+1=0$的伽罗华群的乘法表



当然上面只是伽罗华理论的粗略概述。[3]它对于讨论像F_3或$F_3(a)$这样的置换非常适宜，这两个域分别有三个和九个成员。困扰代数学家好几个世纪的问题是系数在$\mathbb{Q}$中的方程解问题，这是一个有无穷多个成员

的域。如何置换呢？

我希望随着介绍的深入我能把这个问题讲得清楚一些。然而，我怀疑自己能否把事情弄得非常清楚。伽罗华理论是非常难且很精妙的高等代数分支，非数学家不太容易接受。如果你能记住系数属于某个域的多项式可以在一个更大的域里有根，这个更大的域和这个较小域之间的关系可以用群论的语言表示，以及因此每一个与求解多项式方程有关的问题都可以转化成群论中的一个问题，那么你就已经掌握了伽罗华理论的实质。



最后我还要介绍另外一种类型的域，聊表对读者的歉意。在讨论18世纪代数学家的工作时，我一律使用了单词“多项式的”，这仅仅是为了简单起见。其中的一些实际上不是多项式，而是“有理函数”。

有理函数是像下面这样的两个多项式的比：

$$\frac{2x^2+x-3}{3x^3+2x^2-4x-1}$$

通过一些简单的验证可知，任意两个这样的有理函数都可以做加、减、乘和除运算，因此它们形成域。

注意，有理函数域“取决于”另一个域，这个域就是多项式的系数所取的域。事实上，有理函数域可以如上所述那样看成是域的扩张。我从系数域开始。无论它是什么。然后，我加入x并允许加法、减法、乘法和除法都成立。这就生成了有理函数域。这个域与之前域扩张实例的唯一不同是，在那些例子中我对加入的数更了解，知道它的平方或者立方是2，这样我就能对域算术做很多化简。在这里我对x一无所知。它只是一个符号，也可以说它是一个未知量。

注　解

[1] 这不是说在域论中没有更深更难的结果，只是它们不能像群那样对代数方法有直接的帮助，数学家通常是通过代数几何的方法来研究它们的。不过术语

“域论”在数学中有两种不同的意义。在这里的意义是：对被称为“域”的代数对象的研究。另一种意义指的是对某些空间的研究。某些空间是这样的一些空间，在它的每一点都定义了某些量，例如标量或向量，有时候甚至是非常奇怪的量。如果我说“电磁域论”，你就会明白我的意思了。

[2] 事实上，某些教科书的作者喜欢把这个域的元素写成0、1和−1，而不是0、1和2，迈克尔·阿廷就是这样做的。于是这个域的算术看起来就不那么别扭了：1+2=0变成1+(−1)=0。不过还是有(−1)+(−1)=1这种别扭的表达式。

[3] 事实上，这个概述也是事后做了一定改进的版本，算是现代版。伽罗华1830年的原始论文没有使用“域”这个词，爱德华教授的著作《伽罗华理论》中将这篇论文收为附录。直到1879年“域”这一词才有了其代数意义，当时是理查德·戴德金第一次使用它的。顺便提一下，我的例子证明了把某个不可约的二次方程的解添加到F_3中可以构造出F_9。这是一般定理的一个特殊情况：如果$q=p^n$，我们可以通过把F_p的某个n次不可约方程的某个解添加到F_p而构造出F_q。

第 11 章

黎明的枪声

我们不得不承认，数学是一门枯燥的学问，和激情、浪漫不沾边。因此，数学历史学家为伽罗华的故事赋予了更多的传奇色彩。

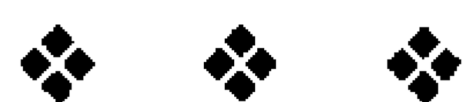
也许是渲染得有点儿过头。根据我们掌握的事实，可以说伽罗华的故事中有很多虚构、错误、推测，甚至街谈巷论的成分。流传最广的英文故事是贝尔的杰作《数学精英》中的一章“天才与笨蛋”。在这一章里，贝尔讲述了这样一个故事，说是一群愚蠢的军人迫害一名充满激情的理想主义天才，天才在他生命中最后一个令人绝望的晚上，完成了成为现代群论基础的论文。贝尔的故事在细节上肯定是虚构的，但是对伽罗华性格的刻画有可能是真的。汤姆·佩茨尼斯1997年的小说《法国数学家》的写作风格不太合我的口味，而且其结尾太荒谬，但是它更接近伽罗华案例的真相。建议浏览宇宙学家、数学史业余爱好者托尼·罗斯曼的网站[1]，这一网站非常客观地评价了所有这方面的资料。

伽罗华在20岁零7个月时死于一场决斗。这场决斗是枪战，发生在黎明。伽罗华显然确信他自己要被杀死，甚至可能希望如此，所以决斗前一天晚上他没有睡觉开始写信。有一些是写给政见相同的朋友，那些像他一样反君主制度的共和党人。在写给他的朋友舍瓦利耶的信中附加了对他的数学著作的解释。

这场决斗的原因不详，可能是政治原因也可能是恋爱原因，或者二者兼有。伽罗华在其中的一封信中写到：“我已经被两位爱国者激怒

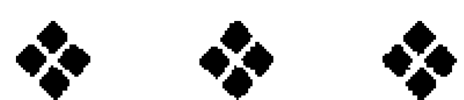
了……。”然而在另一封信中他又说：“我死在了一个荡妇的手里……。”伽罗华卷入了1830年起义前后在巴黎壮大起来的极端共和党政治斗争之中。雨果在《悲惨世界》中描绘了这场起义。他同时因恋爱未果而受到伤害。

毫无疑问，伽罗华的人生是浪漫的。但浪漫的背后往往隐藏着某种病态和不堪。伽罗华的人生也的确充满了悲剧色彩，他自己古怪的性格很大程度上决定了他的不幸，这一点也加剧了其人生的悲情成分。



1830年的法国不是乐土。在拿破仑失败之后，联盟国帮助恢复了波旁王朝，皇帝查尔斯十世年事已高，而且极端保守。社会状况每况愈下，飞速的城市化和工业化把巴黎几乎变成了贫民窟，成千上万的人民生活在水深火热之中。这就是巴尔扎克和雨果笔下的巴黎，整个巴黎冷酷、唯物至上的资产阶级根本不顾愤怒的社会下层的死活。社会下层人民的生计完全受不可控制的经济周期的影响，他们的苦难只有通过有钱人偶尔做做慈善才能略微得到缓解。

1830年，经济不景气。面包价格飞涨，大约有60 000巴黎人没有工作。7月份，设置了路障，暴动分子控制了城市，查尔斯十世被迫逃到了乡下。波旁王朝的远亲奥尔良大公路易·菲利普被进步中产阶级议会代表选为新的国王——7月君主。路易·菲利普为人和蔼谦逊，有几分贵族的自由主义者风格。然而，当时的法国政治日趋激进，纯粹的自由主义无法满足他们。19世纪30年代一直都有起义爆发，其中最重要的一次是1831年发生在巴黎的起义。这是局势紧张的年代，当时，带有强烈信念、头脑发热的年轻人很可能被监视他们的警察抓到，多半要蹲一段时间监狱。



1811年10月，伽罗华出生于巴黎南部的小镇堡拉瑞恩，现在这个小镇位于巴黎通向奥尔良公路的市郊。伽罗华的父亲显然是一名自由主义者、反教条主义者和反保皇主义者。1815年他当选这个小镇的镇长，也就是在拿破仑当皇帝的最后时日、滑铁卢失败前的“百日”期间当选。

君主体制恢复后，老伽罗华发誓忠诚波旁王朝，这不是他的真心话，而只是为了不让真正的保皇派取代他的位置。

我们所知的伽罗华性格的第一份评论来自于他在巴黎私立学校求学时的老师。他们描绘的年轻人是这样的：很聪明但比较内向，工作没有条理，而且也不愿意听从别人的劝告。罗斯曼是这样说的："'出色'、'古怪'、'怪人'、'孤僻'等词语频频出现在伽罗华在路易大帝学校的生活中。连他的家人都开始认为他古怪。"然而，罗斯曼又说伽罗华老师的观点却不同，他的学生时代绝不是贝尔所描述的那样遭受愚钝折磨而非常痛苦。

1829年7月，伽罗华还不到18岁，他的父亲因遭受一位地方牧师的恶意诽谤，在距伽罗华学校只有几码远的一所巴黎公寓内自杀。这一不幸令伽罗华悲痛不已。几天之后，他参加了著名的巴黎理工大学入学考试的面试，其中的考官就有拉格朗日、拉普拉斯、傅里叶和柯西。也许因为他不够老练，也可能因为他故意显示出傲慢自大，伽罗华没有通过考试。当时有这样一幕，考官要求他证明某个数学陈述，他回答说这个陈述应该是很显然的。几个月后，也就是1830年初，伽罗华18岁半，进入了一所没有名望的预备学校，实际上就是一所教师培训学院（现在称为巴黎高等师范学院）。

伽罗华关于方程解理论的第一个版本，也就是贝尔暗指伽罗华在临死的前一天晚上疯狂而草草地写下的那篇论文，实际上在他父亲自杀的几个星期前已经提交到了法国科学院。柯西被指定审阅这篇论文。贝尔认为柯西把这篇论文弄丢了或者根本就置之不理。（公平地说，所有人都这样认为，直到最近法国科学院档案的研究才为柯西平反昭雪。）其实正好相反，这位大人物似乎非常看重这篇论文，他建议伽罗华再仔细润色一下，再为它申报法国科学院数学奖。总之，不管柯西是否这么提议，伽罗华正是那样做的。他于1830年第二次把这篇论文提交给傅里叶，当时法国科学院的秘书。但是傅里叶于5月16日去世了。

也许柯西还能够把伽罗华从黑暗中营救出来。然而，当时正是革命的年代，路易·菲利普的新自由主义政权不符合柯西的胃口。在任何情况下，柯西都是一个很讲原则的人，他已经发誓忠诚于查尔斯十世，就认为不可能再对路易·菲利普发同样的誓言。他本打算辞职，退隐到某个偏远、守旧的地方去（当时他40岁），但最后却背井离乡，离开法国长达八年之久。对于柯西的自行流放没有合理的解释，只有弗罗伊登塔尔（见第7章）所说的“唐吉诃德式行为”的倾向比较符合。

伽罗华自己并没有参加7月革命。巴黎高等师范学院的校长了解到学生当中有很多激进分子之后，把学生们锁了起来，不让他们参加街战。然而，伽罗华很激进，他主动退学了，以享受充分的自由。

这是1831年1月初的事情。伽罗华人生的最后17个月的进程是这样的。

1831年1月4日——从巴黎高等师范学院退学。在接下来的四个月时间里，伽罗华似乎在巴黎当私人数学教师谋生，同时与其他同情共和党的年轻激进分子一起到处游荡。

1月17日——向法国科学院提交关于方程解的论文的第三版。西蒙·丹尼斯·泊松被委派为阅稿人。

5月9日——参加了一群好斗共和党人的宴会。在宴会上，当提议干杯时，他似乎说了威胁路易·菲利普性命的话。第二天伽罗华被捕了。

6月15日——接受了审问，但被无罪释放，可能是因为他年轻的原因。

7月14日（法国国庆日）——伽罗华再次被捕，一同被捕的还有他的朋友欧内斯特，原因是穿了被禁止的炮兵制服。很显然携带枪支也是被捕的原因。据报导说伽罗华带了一把装满子弹的来福枪、“几把手枪”和一把匕首。

（伽罗华从1831年7月14日一直被关押到1832年4月29日。然

而监狱里的条件似乎不是很严酷。例如，犯人经常酗酒。）

10月——接到法国科学院泊松的拒稿信。他发现伽罗华的文章很难理解，但是并没有将其退稿，而是建议他改进其陈述。

1832年3月16日——与同监狱其他人一道被转移到一所疗养院，避免他们染上当时席卷巴黎的霍乱。这所疗养院其实是一所“开放式监狱”，伽罗华可以自由地进进出出。在这里他爱上了斯蒂芬妮·杜姆苔，她是这里一个医生的女儿。但是，他的求爱被拒绝了。

4月29日——伽罗华被释放。

5月14日——伽罗华似乎收到了斯蒂芬妮的拒绝信。

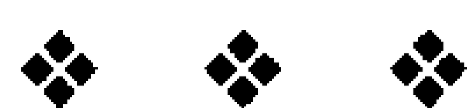
5月25日——伽罗华写了一封信给他的朋友奥古斯特·舍瓦利耶，告诉他自己失恋了。

5月30日——致命的决斗。

这场决斗的确切背景及伽罗华生命最后几天的真实状况尚不为人所知，也可能永远都不会为人所知了。在某种意义上说，极有可能是伽罗华自己放弃了生命。他父亲的死、论文遭到拒绝（还有之前一次投稿的未果）、他对斯蒂芬妮的爱不被接受、就业前景暗淡、几个月的监禁、瘟疫流行的巴黎惨状，这一切对他来说太沉重了。

6月4日，里昂一家报纸对这场决斗做了简要的报导，人们才知道是两个老朋友为了得到某个女人的宠爱而进行决斗，采用俄国式轮盘赌来决定胜负。“这把手枪是对手挑选的，但是因为他们是老朋友，所以彼此不忍看着对方……”这家报纸只确认伽罗华的对手是L.D.。这个女人很可能是斯蒂芬妮，但是谁是L.D.呢？按着当时流行的拼写标准，这个D可能是Duchatelet或者是Perscheux d’Herbinville，或者是其他伽罗华熟悉的共和党人。可是也没有人能确定以上怀疑的人中是否有名字以L开头的人，法国父母对孩子的名字不太在意。

伽罗华的兄弟和朋友抄写了很多份他的论文，寄送给当时的大牌数学家，其中包括高斯，但是没有得到及时回复。最后，在那场决斗的10年后，法国数学家约瑟夫·刘维尔对这些论文产生了兴趣。1843年，他在法国科学院宣读了伽罗华的主要研究成果，并在三年后在一个他自己创办的数学杂志[2]上发表了伽罗华的所有论文。至此，伽罗华的名字才在较大的数学团体中为人所知。



到底是什么使得伽罗华关于方程解的工作对代数方程如此重要呢？在这里我要给出一个简要的描述，但是我使用的是现代语言而不是伽罗华的语言。

在图10-1中，我展示了凯莱表，即三个对象的置换的“乘法表”。一共有六种可能的置换，根据这个表它们可以合成（先做一个置换，然后再做另一个置换）。之后在图10-2中，我给出另外一张表，这是单位的六次方根的乘法表。我曾说过这两张表说明了有六个元素的（即秩为6的）群是什么样的。

这些表展示了抽象群论的本质特征。一个群就是一个对象的集合以及合成这些对象的一个法则，对象可以是置换、数值或任意其他东西。这个法则最具代表性的就是乘法，当然这只是一种方便的记法而已。如果这些对象不是数值，那么这个法则就不是真正的乘法。

为了刻画一个群，集合（对象加上法则）必须满足下面的原理或公理。

- 封闭性。两个元素的合成结果必须是这些对象中的另外一个元素，必须“在这个群内”。
- 结合性。如果a、b、c是这个群里的任意元素，×是合成它们的法则，那么总有$a\times(b\times c)=(a\times b)\times c$。利用这个法则我们能够没有歧义地合成这个群里的三个或更多元素。
- *存在单位元*。在这个群里存在着这样一个元素，当其他任意元素与它合成时都保持不变。如果我们称这个特殊的元素是1，那么

对于这个群中任意元素都有$1 \times a=a$。

- 每个元素都有逆元素。如果a是这个群中的任意元素，那么能够找到一个元素b，满足$b \times a=1$。这个元素就称为元素a的逆元素。通常记为a^{-1}。

这种使用集合的语言“元素”和“合成”、通过公理的手段定义群的高度抽象的方法，就是我已经提到过的20世纪典型的公理方法。伽罗华当然不能使用这种方法，他是用置换群的特殊性质来表述他的想法的。

群中元素的数目称为这个群的秩。很容易就可以验证第10章中两个六元素群满足群的公理。不能如此容易验证的是不存在其他秩为6的群。我指的是抽象群，当然有很多行为如群一样的6个事物的其他实例（我马上会构造一个），但是它们的合成法则都满足第10章的两个凯莱表中的一个。这些乘法表给出了六元素群的两种可能模式。没有其他模式。即只有两种秩为6的抽象群，当然每一种群都有很多实例。在1854年的那篇论文中，凯莱列出了所有秩小于等于6的群。当然今天我们知道的群更多。图11-1给出了秩为n（$1 \leqslant n \leqslant 15$）的抽象群的数目。

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
群的数目	1	1	1	2	1	2	1	5	2	2	1	5	1	2	1

图11-1　秩为n（$1 \leqslant n \leqslant 15$）的群的数目

我们如何能够求得任意给定秩n的群的数目呢？对于这个问题没有一般的方法或公式。然而，还是有一些事实需要注意。例如，如果n是一个素数，似乎秩为n的群不会超过一个。这是正确的。对于任意素数p，秩为p的唯一一个群就是C_p，这个群就是单位1的p次方根且乘法是普通乘法的秩为p的循环群。

下面是两个秩为4的群。它们都是由凯莱发现的。我将用随意的符号α、β、γ表示它们的元素（我将单位元写成1）。

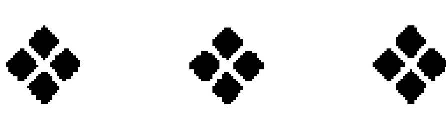
×	1	α	β	γ
1	1	α	β	γ
α	α	β	γ	1
β	β	γ	1	α
γ	γ	1	α	β

×	1	α	β	γ
1	1	α	β	γ
α	α	1	γ	β
β	β	γ	1	α
γ	γ	β	α	1

图11-2　两个秩为4的抽象群，C_4和$C_2\times C_2$

这两个群都有名字。左边群的名字是C_4，是秩为4的循环群，例子就是1的4次方根，令α、β和γ分别等于i、−1和−i时就可以看出这一点。右边群的名字是$C_2\times C_2$，或者称为克莱因4群。它们都是交换群[3]。

观察一下左边的群C_4，如果我告诉你秩为3、5和7的群分别是C_3、C_5和C_7，那么你应该非常容易地写出它们的乘法表。秩为2的唯一一个群就是图FT-3中的群。秩为1的群就不用介绍了，我们把它包含进来只是为了完整起见。因此，现在你知道了所有秩到7的群，比凯莱这个数学家知道的都多了。



伽罗华伟大的洞察力还体现在对抽象群的结构的研究上。观察图11-2中右边的图，这是一个克莱因4群。两个元素1和α在这个群内形成一个小群，一个秩为2的子群。同样，1和β及1和γ也形成子群。如果观察图11-2左边那个群（我称它为C_4），你会看到1和β形成这个群的一个子群，但是1和α以及1和γ却不能形成子群。这就是我说的结构的意思。这些群内群，即*子群*，在群论中担当重要的角色。

回过头来看一下群S_3的乘法表。这个群是三个对象的置换群，如图10-1所示。它有若干个子群。有一个秩为3的子群是由I、(123)和(132)组成的（请注意它们都是偶置换）。它还有3个秩为2的子群，一个是由I和(23)组成的，一个是由I和(13)组成的，最后一个是由I和(12)组成的。这4个子群的每一个都是一个圆满的、自给自足的单元。在这里面可以随意做乘法，随意取逆，所得结果都在这个子群内。在第二个称为C_6的秩为6的子

群（见图10-2）中，有一个秩为3的子群，是由1的3次方根组成的$(1,\omega,\omega^2)$，还有一个秩为2的子群，是由1的平方根$(1,-1)$组成的。

群结构的第一个伟大定理是拉格朗日定理，我已在第7章中提到过了：子群的秩整除这个群的秩。这个除法的商叫做这个子群的指数。拉格朗日定理排除了小数指数。我们可以在秩为6的群内寻找秩为2或3（指数为3或者2）的子群。但是我们不可能找到秩为4或5的子群，因为4和5都不能被6整除。

伽罗华为群结构引入了一个关键的概念，这个概念就是今天所说的*正规子群*。下面对此给予简要的说明。以S_3和由I和(12)形成的子群为例，我们把这个子群记为H，它是一个秩为2的抽象群，如图FT-3所示。

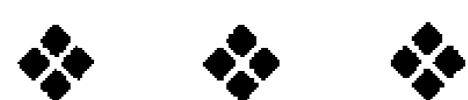
从主群中取出一个元素，这个元素或者在这个子群内或者在这个子群外，这无关紧要。我使用(123)。依次用这个元素与I和(12)相乘。从左到右做乘法，先运用(123)，然后再运用另一个置换，如第10章所述。结果得到一个集合，而不是一个子群，这个集合是由(123)和(23)组成的。这叫做H的“左陪集”。让主群S_3的其他每一个元素重复刚刚用(123)做的运算。这样就会得到一个左陪集的族。其中一个元素就是我已给出的左陪集{(123), (23)}。（注意：大括号是表示集合的通用方法。伦敦、巴黎和罗马组成的集合的数学表示是{伦敦, 巴黎, 罗马}。）此外，有一个是{(132), (13)}，还有一个是{I, (12)}，这就是H本身。（因为S_3的秩是6，所以你也许希望在左陪集中有六个元素，但是实际上它们两两相同。）

重复一下整个过程，但是这次的乘法是“从右边开始”。结果得到一个“右陪集”族：{(123), (13)}、{(132), (23)}和{I, (12)}。如果左陪集族等于右陪集族，那么H就是一个正规子群。在我的例子中，这两个族是不相等的，所以这个特殊的H不是S_3的一个正规子群，只是一个普通的子群而已。然而，由所有偶置换组成的秩为3的S_3的子群是一个正规子群。我把它留给读者自己去验证。按着上面的定义，注意下面的事实：如果一个群是可交换的，那么每一个子群都是正规的。

伽罗华证明，对于一个未知量的次数为n的任意多项式方程

$$x^n + px^{n-1} + qx^{n-2} + \cdots = 0$$

通过研究系数域与解域之间的关系，我们可以将其求解过程同群挂勾。如果这个方程的“伽罗华群”有满足特定条件的结构，其中正规子群的概念最为重要，那么我们就能够仅用加法、减法、乘法、除法和方根表示这个方程的解。如果它没有这种结构，那么我们就不能这样表示这个方程的解。如果n小于5，这个方程的伽罗华群总是有相应的结构。如果n大于等于5，那么它可能有也可能没有这样的结构，这主要取决于p、q及其他系数的数值。伽罗华揭示了这个群结构的条件，因此对“什么情况下有可能找到多项式方程的代数解”这个问题给出了最终的明确答案。



虽然伽罗华的工作标志着方程故事的结束，但是它也标志着群故事的开始。因此本书对伽罗华理论的定位就是：它是开始，不是结束。

我在上文中放下了刘维尔于1846年出版了伽罗华的论文的那条线索。于是，如同现在一样，人们可以认为伽罗华理论是一个结束，也可以认为这是一个开始。注意到伽罗华的工作的数学家大多数都接受它是一个“终结”的观点。历时几个世纪的一些问题大都与多项式方程有关，这些问题已经被彻底解决了。好！现在让我们投身于新的、更有前途的数学领域：函数理论、非欧几何、四元数等。

面向未来取得第一个真正意义上的转变的数学家是同阿贝尔一样的一位挪威数学家。这个人就是路德维希·西罗，他于1832年出生在奥斯陆，这一年伽罗华去世。

挪威数学家远远多于本地的需求，所以，西罗的职业生涯几乎都是在哈尔登镇当一名高中数学教师，这个镇在奥斯陆以东80公里，当时叫弗莱德瑞克绍德。直到60多岁，他才得到大学的职位。在中学教书这些年中他一直坚持自己的数学研究。

西罗很自然地被他的同胞阿贝尔关于方程可解性的工作所吸引。大

约在19世纪50年代末，克里斯蒂安娜大学的一名教授给他看了伽罗华的论文，然后西罗开始研究置换群。尽管凯莱1854年已经发表了论文，但请记住，抽象群论仍然不是数学家关注的焦点。群论还只是一个置换的理论，其仅有的成果就是对代数方程解的研究。

1861年，西罗得到政府奖学金去欧洲其他国家游学一年。他参观了巴黎和柏林，参加了当时几位数学名人的讲座，其中包括15年前拯救并出版了伽罗华论文的刘维尔。返回奥斯陆后，西罗在大学举办了关于置换群的讲座。这是19世纪70年代关于群论的少有的讲座之一，[4]它之所以有趣还有另外的原因，我将在第13章中介绍。

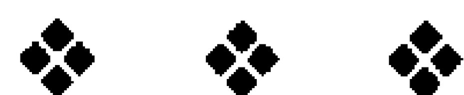
西罗的研究涉及了置换群的结构。我已经提到了拉格朗日定理，这一定理给出了H是G的子群的必要条件：H的秩必须整除G的秩。因此秩为6的群可能有秩为2或3的子群，但是不能有秩为4和5的子群。

西罗的研究集中在“可能”这个词上。的确，秩为6的群可能有秩为2或3的子群。但是它真有吗？对子群来说，我们能够得到比拉格朗日的必要但非充分的整除条件更好的法则吗？柯西已经证明，如果一个群的秩有素数因子p，那么存在唯一一个秩为p的子群。这个结果还能改进吗？

当然能。在1872年发表的一篇论文中，西罗对这一论题给出了三个定理。直到今天这三个定理仍作为群论的基础结果教授给学生。我只给出第一个定理。

> *西罗第一定理。假设G是一个秩为n的群，p是n的一个素数因子，且k是满足p^k是整除n的p的最大幂。[例如：$n=24, p=2, k=3$。]那么G有一个秩为p^k的子群。*

这种秩为p^k的子群称为G的西罗p子群。曾经在世界上某个大学的数学系就有个摇滚乐队，自称“西罗和他的p子群”。



有限群论本身就有着悠久迷人的历史。它还在市场调研到宇宙研究

的若干领域有广泛的应用。完整的群分类也得到了发展，它把不同秩的群分类到不同的族。

我在第10章中给出的乘法表的两个群就是这些族中的两个例子。第一个是S_3，这个群是三个对象所有可能置换的群，它的秩是6。第二个是C_6，这个群是由1的六次方根组成的，秩也是6。它们都是群的族的成员。n个对象的所有可能置换的集合为群S_n，其中置换的合成方法按一般方法，这个群就是秩为$n!$（即n的阶乘）的*对称群*。1的n次方根组成群C_n，其合成方法按一般乘法，这个群就是秩为n的*循环群*。事实上，我们已经发现了第三个重要的群族：从S_n中提取偶置换组成正规子群，这个正规子群称作秩为$(1/2)n!$的*交替群*，或更简单地说是“指数为2的交替群”，总是记作A_n。

另一个重要的族就是*二面体群*。“二面体”的意思是几何概念中的“有两个面”，并不是指人的两面性。从一张硬纸板上切下一个正方形。这个正方形有两个面，它是二面的。按着A、B、C和D这样的顺序标出它的四个角。把这个正方形放在一张纸上，然后用铅笔画出它的轮廓。请问，我能够用多少种基本的初等方法（所有其他方法都与它们等价，例如，720度旋转等价于360度旋转）移动这个正方形，使得它总是能够安全而且精确地与起始的轮廓吻合？

答案是有8种方法。在图11-3中我绘制出了这些方法，而且给出了对特定起始图形作用后的效果，这个起始图形用不做任何改变的恒等移动表示。所有的移动方法如下：恒等移动，顺时针旋转90°、180°和270°的旋转（3种），根据翻转轴不同而不同的翻转（4种）。（南北向、东西向、西南到东北向，东南到西北向，它们就是这个正方形的四个翻转轴。）

这些移动（对数学家来说就是“变换”）在下面的合成法则下形成一个群：先对第一个做基本移动，然后再对另一个做基本移动。显然这个群的秩是8。它的名字是D_4，是这个正方形的二面体群。你可以尝试着写出这个群的凯莱表。对应于任意正n边多边形有相应的一个群，称其为

D_n。当n=2时，这个“多边形”只是线段AB，D_2只有两个成员，但是对于每个大于等于3的n，D_n有2n个成员。所以，这是另外一个群族。

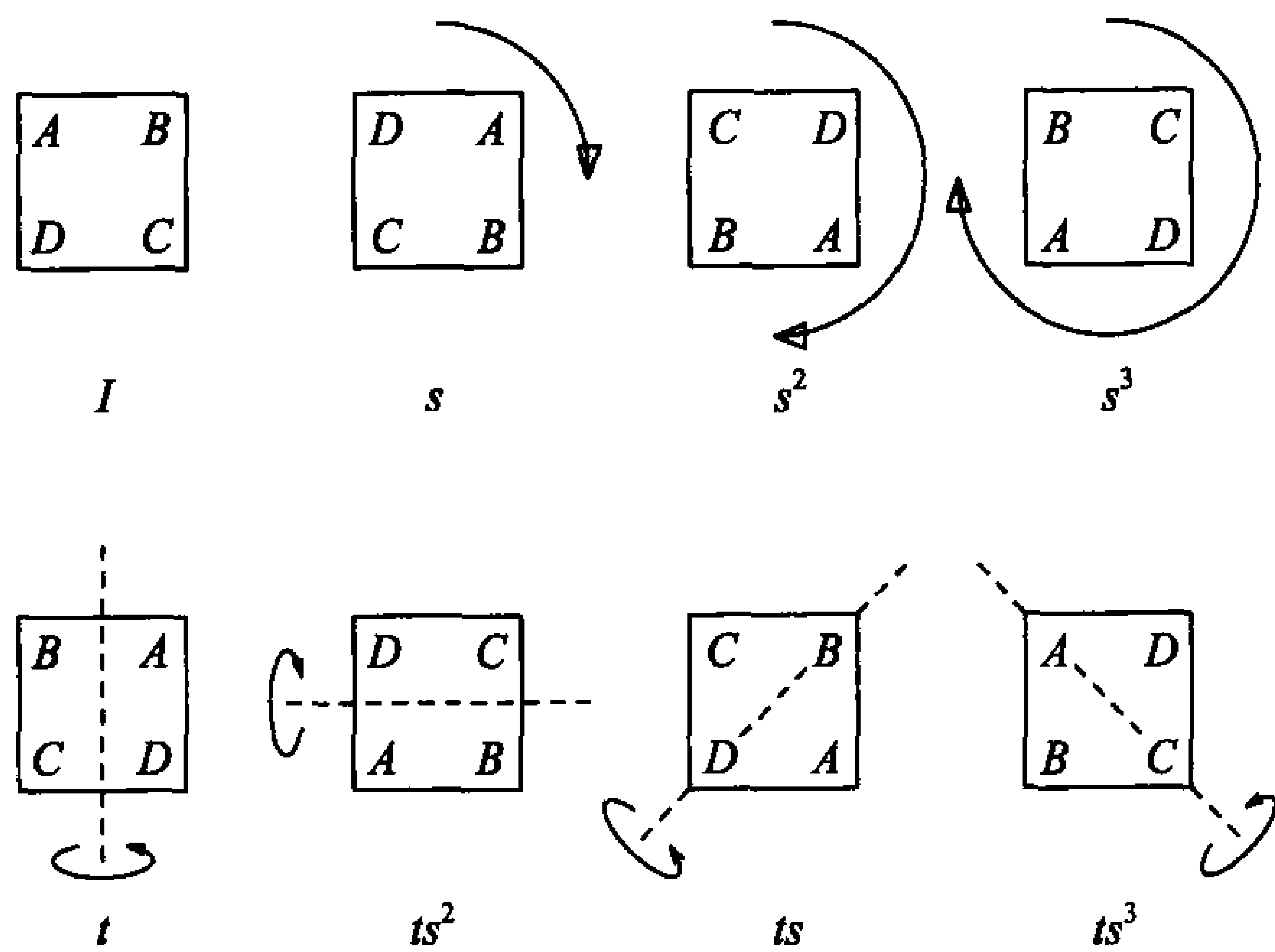


图11-3　二面体群D_4的8个元素

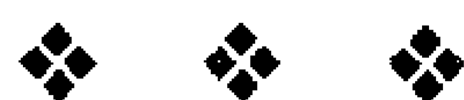
（注意我对D_4成员的有些缜密的记法。例如，一旦我定义90°旋转为s，合成法则是“先做第一个变换，再做第二个变换”，那么180°旋转就是s^2，做一个s再做另外一个s。270°旋转是s^3。类似地，我只需定义南北轴的翻转为t，于是，西南到东北向的翻转就是ts，等等。当你能够用最少的基本元素（如s和t）构建一个群时，这些元素就称为这个群的生成元。图11-2中的秩为4的群的生成元是什么？）

假设我用三角形做这个练习，得到D_3。D_3会有6个成员吗？我不是已经说过只有两个秩为6的抽象群S_3、C_6了吗？这个小难题的答案是D_3是S_3的一个特例[5]。如果考虑到置换，你就会明白为什么对三角形答案是这样的，而对正方形答案就不对了，而且对超过三条边的任意正多边形答案都不是这样的。A、B和C的任意置换都对应于D_3中的二面体的一个动作。对于A、B、C和D则不是如此。置换（AC）对应于动作ts，但是，置换（AB）却不与D_4中的任意动作对应。（另外，同S_3一样，它也有正规子

群和普通子群，你可以尝试着把它们列出来。）

我已经提过，如果p是一个素数，秩为p的唯一群是以1的p次方根为模型的循环群C_p。而如果p是一个大于2的素数，则只有两个秩为$2p$的群，其中一个是C_{2p}，另一个是D_p。

回过头来看图11-1，这个图给出了每个秩的群的数目，现在，你基本可以判断出n=11的情况。美中不足的是n=8。n=8时，有5个不同的群，其中3个是循环群及从循环群建立起来的群：C_8、$C_4 \times C_2$和$C_2 \times C_2 \times C_2$。还有一个当然是$D_4$，第五个是古怪的四元数群，这个四元数群有一个非常特殊的性质：虽然它是不可交换的，但它的所有子群都是正规的。



到此，我希望你能够理解对群进行分类的迷人之处。这个领域中最具历史趣味性的领域也许就是对所有简单有限群进行分类的尝试。简单群是没有正规子群的群。例如，对于素数p，C_p没有真子群，所以默认就是简单群。当n等于5或者更大时，n个对象的偶置换群A_n是简单的，这就是为什么一般五次方程没有代数解的原因。还有另外5个简单群族，以及26个"怪物"——不属于任何族的"一次性"简单群。其中一个"怪物"的秩是

808 017 424 794 512 875 886 459 904 961 710 757 005 754 368 000 000 000

所有简单有限群的分类[6]于20世纪中后期实现，并于1980年最终完成。这是现代代数的伟大成就之一。

注　解

[1] 网址是dilip.chem.wfu.edu/Rothman/galois.html。

[2] 这本杂志是《纯数学与应用数学杂志》，创办初期的名字是《刘维尔的杂志》。该杂志创办于1836年，至今仍有强大的影响力，自称是"世界第二古老的数学杂志"，第一古老的杂志是克雷尔创办于1826年的杂志。

[3] 我不知怎么忘记说明了现在把交换群称为阿贝尔群，以纪念阿贝尔的一个定理。有一个老掉牙的数学笑话："问题：什么东西是紫色的并且可以交换？

答：一个阿贝尔葡萄。”在研究阿贝尔群的时候，用加法而不是通常使用的乘法表示它的运算。因此，阿贝尔群的恒等元通常用0表示（因为$0+a=a$永远成立），而元素a的逆元写成$-a$。为简单起见，下文我将忽略这些。

[4] 19世纪50年代后期，戴德金在哥廷根开设了几次关于伽罗华理论的讲座。

[5] 更精确地说，D_3和S_3二者都属于同一个抽象群。唯一秩为2的抽象群不仅有D_2和S_2，还有C_2，如图FT-3所示。严格说来，所有诸如D_3、S_3、C_2的记法表示抽象群的特殊例子，我们应该避免“这个群S_3”这样的说法，而应该说“在所有常见的群中，S_3是我们最熟悉的例子”，但是也没有人费事去这样严格地叙述。

[6] 我不再详述有限群的分类问题。要想看到更全面、更清晰的记述，请参考基思·德夫林1999年所写的《数学，新黄金年代》一书。如果想看一下最终的分类情况，请参考康威等人编写、由牛津大学出版社出版的《有限群图表集》，只是其表述更专业。

第12章 环小姐

仅通过封闭、结合、恒等元和逆元公理这4个公理就可以定义群（见第11章）。这种简单的说法扩大了群的范围，这就像“四条腿的生物”要比“四条腿且有獠牙和嘴的生物”更多一样。正是由于群定义的这种简洁性，使得群的适用领域不仅限于纯数值，也使得有正规子群这一重要概念的群拥有了复杂而有趣的内部结构。

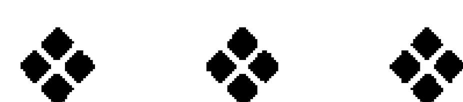
域是一个更复杂的对象，它的定义需要10条公理。它的基本合成法则不只一个，而是两个：加和乘。［减和除分别是加和乘的逆：8−3等于8+(−3)。］这种更高的复杂性使得“域”这一概念与普通的数值关联更加密切。同时这种复杂的定义限制了它拥有有趣内部结构的可能性。

现代代数学研究较多的还有一个数学对象：环。环比群复杂，但是没有域复杂，因此尽管它的应用范围不如群广泛，但是它对普通数值之外的应用比域要广。同群类似，环可以有有趣的内部结构。它的关键概念是理想。我将在这一章和下一章详细介绍。

事实上，同一些中间概念的情况一样，环的概念能够为数学家同时提供两个世界中某些最好的东西。例如，它是现代代数几何的中心，这是现代代数最深奥、最具挑战性的思想的源头。然而，经过了漫长的时间后人们才开始重视环如此强大的力量。

人们常说环论始于费马大定理，这一说法实际上是错误的。然而，费马大定理是一个不错的切入点，它引出了环论，因此，我从这里开始

陈述。



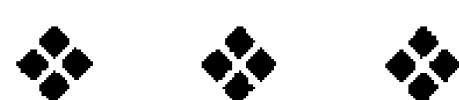
我曾经在第2章提到过费马和他的大定理。费马把这个定理写在了丢番图1637年的著作《算术》一书的页边空白处。这个定理说当n大于2时，方程

$$x^n + y^n = z^n$$

没有正整数解x、y、z。1659年费马自己对于n=4的情况给出了一个大致的证明，后来高斯给出了更完整的证明。1753年欧拉给出了n=3时的证明。

此后，直到法国数学家索菲·热尔曼之前的半个世纪对此都没有什么实质性的进展。热尔曼是本书介绍的三位女性数学家中的第二位（海帕希亚是第一位），她证明费马大定理对于更大一类整数x、y、z和n为真。解释什么样的四个整数符合那类整数恐怕会离题太远，只要阐述这个定理在热尔曼的结果之后的进展就足够了。

法国数学家阿德利昂·玛利埃·勒让德（当时72岁）和德国人狄利克雷（他有一个法式名字）分别证明了这个定理在n=5时的情况。到此，每一个人都知道真正的挑战是对素数n证明这一定理，下一个目标是n=7。另一个法国人拉梅于1839年证明了n=7时的情况。然而，这时，故事出现了新的转折。



提醒一下，整数集合$\mathbb{Z}$是由所有正整数、所有负整数和零组成的。

$$\mathbb{Z}: \cdots -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, \cdots$$

（这个列表在左右两端无限延伸。）这里，$\mathbb{Z}$不是域。任意数之间的加、减和乘运算的结果仍在$\mathbb{Z}$中。但是，除法只在某些时候才可行。如果用4去除−12，那么结果仍在$\mathbb{Z}$中。如果用7去除−12，那么结果不在$\mathbb{Z}$中。因为不能在$\mathbb{Z}$内自由地做除法，所以它不是一个域。

但是，$\mathbb{Z}$本身非常有意思，而且很重要，值得数学家关注。即使除法不可行，还可以研究加法、减法和乘法。例如，可以研究因数分解和

素数问题（即什么样的数是素数）。

另外，还有很多其他数学对象与 ℤ 类似，可以自由地做加、减和乘，但不能自由地做除法。例如，多项式就是如此。给定两个多项式，如x^5-x和$2x^2+3x+1$，可以把它们加到一起得到一个多项式（结果是x^5+2x^2+2x+1），或者把它们相减（结果是x^5-2x^2-4x-1），或者把它们相乘（结果是$2x^7+3x^6+x^5-2x^3-3x^2-x$）。但把它们相除不一定能够得到一个多项式。对于上面这一特殊情况我们就不能做除法，但在有些情况下又可以这样做，例如$(2x^2+3x+1)\div(x+1)=(2x+1)$。就像整数一样！[1]

这种可以进行前三种计算、但不能进行第四种计算的数学对象称为环[2]。现在你就明白了我所说的环介于群和域之间的意思了。“域”的定义比环更严格，可以在其中做除法。“群”的定义更宽松，只有一种合成两个元素的方法。参看第11章定义抽象群的4个公理。域需要10条公理，环只需要6条公理。

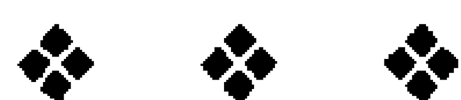
在有关数论的某些研究中，高斯发现了包括复数在内的一类新环。当时高斯不可能使用环这一词，一百年之后它才露面，也不可能使用环的抽象意义，不过他发现的的确是环。这个环就是现在我们称为高斯整数的环，就是由像(−17+22i)这样实数部分和虚数部分都属于整数 ℤ 的复数组成的环。你也可以像高斯一样对这样的“复整数”开发类似于整数的算术。

这种算术并非一目了然。一般说来，环不只是不能自由地做除法。但不难在 ℤ 中引入部分除法，并从普通算术出发开发类似的素数理论和因数分解理论。负数包含在 ℤ 中让事情更棘手了。建立良好的素数和因数分解理论并把其安插进其他环中通常更加困难。对于高斯整数这是可行的。对于欧拉在证明$n=3$的费马大定理时所使用的环，这也是可行的，这个环比高斯整数的范围更广，在允许整数的同时还允许$\sqrt{3}/2$这样的数存在。但是，当你更深入地研究这样的环时，就会遇到麻烦事。

最麻烦的事情就是因数分解唯一性的瓦解。在环 ℤ 中，任意的整数

都只能用一种方法表示成一个单位元与一组素数之积。（在环论中，单位元就是能够分解1的数。$\mathbb{Z}$ 有两个单位元1和−1。复整数的高斯环有四个单位元：1、−1、i和−i。[3]）例如，整数−28分解成−1×2×2×7。除了改变这些因子的顺序之外，无法得出其他不同的因数分解。环 $\mathbb{Z}$ 具有保持因数分解唯一性的性质。

另一方面，考虑数$a+b\sqrt{5}\,i$的环，其中a和b是一般整数。在这个环中，数值6有两种因数分解方法，一种是2×3，一种是$(1+\sqrt{5}i)\times(1-\sqrt{5}i)$。注意了，在这个环中的四个因子都是素数（素数的定义是除了单位元和它们自己之外没有其他因子）。唯一因数分解性瓦解了。



因数分解唯一性不成立引发了1847年数学界的动荡。当时，法国科学院为费马大定理的证明设立了金牌和3000法朗的奖金。在成功证明了n=7时的情况之后，那一年的3月1日，拉梅被通知参加法国科学院会议，因为他接近完成这个定理的一般证明，即对于所有n的证明。他补充说，这一证明的想法在几个月前同刘维尔交流时就有了。

当拉梅完成证明时，刘维尔本人站了出来，对拉梅的方法泼了冷水。他指出：第一，这个证明不是原创的；第二，这个证明依赖在特定复数环内因数分解的唯一性，这一性质是不可靠的。

柯西发言了。他支持拉梅，说拉梅的方法完全可能生成一个证明，并透露他本人也在沿着相同的思路进行研究，很快就能给出他自己的证明。

这次科学院会议之后，人们对费马大定理疯狂研究了几个星期，其中不但有拉梅、柯西等人，还有其他被奖金吸引的人。

接着，12周之后，在拉梅和柯西宣称他们完成了证明之前，刘维尔在科学院宣读了一封信。这封信来自德国的一位数学家库默尔，他看到了关于这个数学传言的巴黎公报。库默尔指出利用拉梅和柯西所采用的方法，因数分解的唯一性的确不成立，并声称他自己在三年前已经证明

了这个结果（尽管他是在一本知名度非常低的杂志上发表他的证明的），还说如果使用他一年前发表的一个概念，那么这种状况就会得到某种程度的改善。这个概念就是理想因子。

据说（是贝尔在《数学精英》一书中说的），库默尔在他证明费马大定理时发明了这一新概念。然而，现代学者认为库默尔在发现这些理想因子之后对这一定理没有做什么工作。只是在向巴黎的那些大惊小怪的人们发出警告后，他才开始挑战这一定理。

在刘维尔读这封信的几周后，库默尔向柏林科学院提交了一篇论文。在这篇论文中他对一大类素数，即所谓的正则素数[4]证明了费马大定理。他的证明使用了他发现的理想因子。这是一个世纪以来在攻克费马大定理的征程中所取得的最后的重要进展。这个定理最终由安德鲁·怀尔斯于1994年证明。

而这些理想因子是什么呢？这不太好解释。实际上，数学史学家通常不去解释它，因为不久库默尔的理想因子就被更强大更一般的理想的概念所取代。理想不是一个数值，而是一个数值的环。我认为这对库默尔有点不公平，所以在这里大概介绍一下他的概念。

库默尔利用了割圆整数这一概念，我先简要说明这一概念。读者也许回想起介绍单位根时提到的“割圆”。当这个词在数学中出现时，通常都在单位根范畴内。假设p是某个素数，单位的p次方根是什么？好的，当然1是单位的一个p次方根。单位的其他p次方根都均匀分布在复平面内单位圆的圆周上，如图RU-1所示。如果称第一个根（从1开始顺时针方向前进）为α，那么其他的根是α^2、α^3、α^4、…、α^{p-1}。

割圆整数是如下形式的复数

$$A+B\alpha+C\alpha^2+\cdots+K\alpha^{p-1}$$

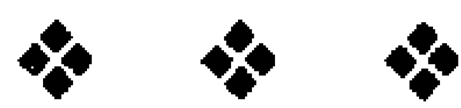
其中所有大写字母系数都是$\mathbb{Z}$中的普通整数，α是单位的一个p次方根。例如，如果p是3，那么这些单位根就是多次提到的1、ω和ω^2，后面两个是二次方程$1+\omega+\omega^2=0$的根（见“数学知识：单位根”）。对于$p=3$的情况，

一个割圆整数的例子就是$(7-15\omega+2\omega^2)$。注意，这是一个非常普通的复数$[27/2-(17/2)\sqrt{3}\,i]$。而我刚才是用1、ω和ω^2表示它的。

这些割圆整数有一些神奇美妙的性质。仍以$p=3$为例，因为$1+\omega+\omega^2=0$，所以对于整数n，有$n+n\omega+n\omega^2=0$。因为加零保持一个数不变，我可以在左边加上$(7-15\omega+2\omega^2)$，得到$[(n+7)+(n-15)\omega+(n+2)\omega^2]$，没有改变这个数，你可以代入ω和$\omega^2$的真实值验证。正如3/4、6/8、15/20、75/100等无穷多个有理数表示同一个有理数一样，对于n的任意值，我的割圆整数的第二个形式都将表示同一个割圆整数。

好了，库默尔的工作就是关于这些割圆整数的因数分解。事实证明，这是一个很深奥的棘手问题。是的，因数分解唯一性瓦解的问题很快又出现了。（然而也不是非常快：当$p=23$时才首次出现这样的问题。这就是理想因子理论很难说明的一个原因。）这就是库默尔解决的一个特殊问题。他把素数的普通定义加以限制使其更适合割圆整数，从而解决了这个问题。于是库默尔从“真素数”出发创建了他的理想因子，从而得到了割圆整数因数分解的完整理论。

除此之外，库默尔还证明了他的伟大结果：费马大定理对于正规素数成立。但是，这只是特殊的、范围有限的应用。在环论的强大得到充分展示之前，人们获得了更高层次的一般化。这一更高层次是经过下一代数学家的努力达到的。



库默尔于1847年给法国科学院的信的意义已超越了数学范畴。库默尔当时37岁，是普鲁士布雷斯劳大学的一名教授。虽然德国的统一是20年之后的事情，但是民族情绪很强烈，即便不是作为一个国家而是作为一个民族，德国人在欧洲文化中也是一股越来越强大的力量。拿破仑战争期间因法国对德国的轻蔑而产生的愤恨一直折磨着德国人，这样的折磨战后一直持续了40年。

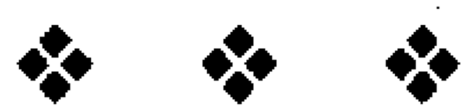
库默尔对这种愤恨有着切身的体会。他的父亲是距柏林东南160公里

叫索劳[5]的小镇的一名医生，在库默尔三岁时，他因拿破仑大军从俄罗斯撤退时带到这个地区的伤寒而去世。因此，库默尔是在极度贫穷中长大的。尽管他似乎是一位讨人喜欢的小伙子，是一位有才华的老师，但是这一切不禁让人猜测当法国科学在得意地显示它的大佬地位时，库默尔的内心一定会感到阵阵刺痛。

拿破仑给德国人造成的挫败感和耻辱感也带来了更大的影响。首先德国人加速普鲁士乃至德国诸小州的建设，加速教育、技术和教师培训体制的改革。这一改革以及伟人高斯的榜样和威望造就了中世纪一批一流的德国数学家：狄里克雷、库默尔、亥姆赫兹、克罗内克、艾森斯坦、黎曼、戴德金和克莱布施。

到了1866年统一的时候，德国甚至已经有了两个值得夸耀的数学中心：柏林和哥廷根，它们有各自的独特风格。柏林数学崇尚纯净、稳重和严谨；[6]而哥廷根数学则更富于想象，且倾向于几何学。如果把前者比作罗马，后者就是雅典。魏尔斯特拉斯和黎曼就是两种风格的典型代表。如果没有所需的长达八页的细心证明，柏林学派的魏尔斯特拉斯不会理会什么理论。另一方面，黎曼抛出了漫游于复平面、弯曲表面以及自相交表面的惊人函数，偶尔停下来为所需的草案给出一个草草的证明。

而与此同时，法国的数学已开始衰退。这样说也是相对的：这个国家有值得夸耀的刘维尔、埃尔米特、伯特兰、马修和乔丹，她不缺乏数学天才。然而，巴黎数学辉煌的时代已成为过去。柯西在高斯去世两年后也撒手人寰了。柯西的去世标志着法国数学伟大时代的结束，而高斯的去世正是发生在德国数学快速崛起的时候。



戴德金是德国中世纪盛产的一批数学家中最出色的一位。他是一个非常安静、沉默寡言的人，除了数学之外对任何事情都不关心。戴德金的一生波澜不惊，大部分时间都是在他的家乡（也是高斯的家乡）布伦

瑞克的一所大学当教师。

戴德金对代数的贡献有三方面。第一，他给出了理想的概念。第二，他与韦伯开启了函数域理论，关于这方面的内容我已经在关于域的知识的结尾给出了简单的介绍。（在第13章中将有更详细的介绍。）第三，戴德金开启了代数公理化的进程，把代数对象作为纯抽象对象，利用集合语言的方式进行定义。半个世纪后，当这种公理化方法充分成熟时，它将变成现代代数观点的基础。

理想的概念不是一个能够让非数学家容易理解的概念，因为不太容易举出启发性的例子。首先，一个理想是一个子环，是一个环里面的环。因此，它是一个数（或多项式，或组成父环的其他对象）的族，在加法、减法和乘法下封闭，嵌入在相同类型的更大的族之内。

但是，理想不仅仅是子环。它有这样的特性：如果你任意取它的一个元素，把这个元素与大环中的一个元素相乘，其结果仍局限在这个子环里面。

取 $\mathbb{Z}$ 这个我们最熟悉的环，$\mathbb{Z}$ 中一个理想的例子是：某个给定数的所有整数倍的全体。假设我们取的这个数是15。下面就是一个理想：

$$\cdots, -60, -45, -30, -15, 0, 15, 30, 45, 60, 75, \cdots$$

这个理想是所有形如$15m$的整数组成的，其中m是任意整数。简言之，理想在加、减和乘运算下是封闭的。而且，正如我预告的那样，如果你把上面这个理想中的任意一个数，比如30，乘以大环 $\mathbb{Z}$ 中的任意一个数，比如2，结果60还在这个理想之中。

如果我用下面的方法对这个例子进行扩展，也许会更好。现在我从 $\mathbb{Z}$ 中任意取两个数，构造它们的所有线性组合。比如，我取两个数15和22，构造它们的所有线性组合$(15m+22n)$，其中m和n是任意整数。这也许是一个更有趣的理想。

但使用 $\mathbb{Z}$ 进行扩展时，不会增加新的理想，因为 $\mathbb{Z}$ 的结构太简单。如果你让m和n跑遍整个 $\mathbb{Z}$ ，那么$(15m+22n)$取遍所有可能的整数，这很

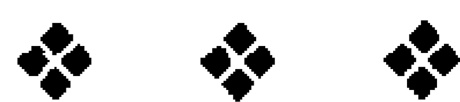
容易证明。[7]所以，这个“理想”实际上就是整个$\mathbb{Z}$。如果不取15和22，而是取有公因子的两个数，比如说15和21，我们将得到由它们的最大公因数3生成的理想。所以，上面给出的这种理想是$\mathbb{Z}$中除$\mathbb{Z}$本身（和只由零组成的平凡理想）之外仅有的理想。事实上，$\mathbb{Z}$中的理想不是非常有趣。

下面介绍用形式化抽象代数的语言进行描述的方法。在任意环中，某特殊元素a的倍数的集合称为由元素a生成的**主理想**。像$\mathbb{Z}$这样每一个理想都是主理想的环，称为**主理想环**。在一个不是主理想环的环中，你可以这样生成理想：取两个或更多的元素a、b等，构造它们的所有可能组合$(am+bn+\cdots)$，这个理想被称为由“$a,b\cdots$生成的理想”。其实，分类环的一个方法就是审查这个环的理想的生成方法。例如，存在一种被称为诺特环的重要类型的环，它的所有理想都是由有限个元素生成的。

在复数环中，理想则变得非常有趣。戴德金给出理想的抽象定义，这就是刚才我给出的，并把这样的定义运用于许多种类的复数环中，这一类环要比库默尔使用的环范围广得多。这样做，他就能够生成适合任意环的“素数”、“约数”、“倍数”和“因子”等概念的定义。

戴德金用比以往数学家所尝试的更一般的方法表述这些定义。他没有完全使自己脱离开数值的领域，但是他引入了他的数学对象——域、环（他称其为“序”）、理想、模（这是一个向量空间，其标量取自于环，而不是域），并定义公理，就如现代代数课本里所做的那样。因为他还没有现代集合论的术语可用，所以戴德金的定义看起来还不够现代，但是他的路线是正确的。

在下一章介绍代数几何时，我将更详细地讨论理想。



一旦戴德金的方法传播开来并被人们接受，理想的概念为人们所熟悉之后，环有像群一样的有趣内部结构的事实也就变得清晰了。那时就是环论起飞的时候。然而，戴德金的方法还称不上是环论。使用它的人

总是在几何、分析、数论中，特别是在代数中找到它的应用。这里的代数特别指的是多项式。直到第一次世界大战之后，我们的环小姐才款款登场，带来了相应的理论，该理论涵盖了所有这些领域，把它们巩固在稳固的公理定义之下。

我下一节再介绍这位小姐。在戴德金与这位小姐之间的40年空隙里，这一理论被很多数学家向前推进，其中包括一些伟大的数学家。然而，他们的努力最有意义的是在几何方面，属于下一章的内容。这里，我将只提及那一时期环论中的一个名字，因为这个人及他的生活确实很有意思。

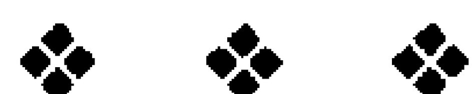
这个名字就是伊曼纽尔·拉斯克，人们知道他不是因为他是数学家，而是一名棋手。事实上，他是连续27年的国际象棋冠军，从1894年一直到1921年，他是拥有这一头衔时间最长的人。

1868年拉斯克出生在当时属于德国东部的一个地区，第二次世界大战之后，边界重新划分，这一地区成为波兰西部的一部分。他是犹太人，他的父亲是他们小镇的犹太人教堂的合唱指挥，当时这个小镇称为柏林宸，现在称为巴尔利内克。拉斯克跟他大哥学习国际象棋，到十几岁时就在镇上的咖啡屋下国际象棋挣小钱。他的棋艺进步很快，20岁时就赢得了他的第一次锦标赛冠军，25岁时在北美（纽约、费城、蒙特利尔）举行的一系列比赛中打败了占据冠军宝座的威廉·斯坦尼兹，成为世界冠军。

拉斯克的数学教育非常全面，但是被他的国际象棋活动所扰乱。在进入柏林、哥廷根、海德堡等地的大学后，从1900年到1902年他在埃尔朗根大学师从希尔伯特，33岁时在那里获得博士学位。他对环论的主要贡献是确立了相当深奥的准素理想的概念，这个概念有点像因数分解一个整数时的素数的幂（例如$6776=2^3\times7\times11^2$）。有一种类型的环称为拉斯克环，还有一个重要定理是拉斯克–诺特定理，这是一个关于诺特环结构的定理。

拉斯克老年很凄惨。1933年希特勒掌握政权时，他和他的妻子已在

德国安定下来过着舒适的退休生活。纳粹没收了拉斯克的所有财产，把他们弄得一贫如洗并撵出了故乡。因此在60多岁的时候，他不得不重新参加国际象棋比赛。他在英格兰生活了两年，然后去了莫斯科，最后又辗转到了纽约，1941年在那里去世[8]。



诺特环和拉斯克-诺特定理表明，在这个故事中显然有一个叫诺特的人。实际上有两个人，一个贡献小些，一个贡献更大。贡献小的是父亲马克斯·诺特，贡献大的是他的女儿爱米·诺特。戴德金完成他的开创性工作之后的40年间还涌现了很多成果，是爱米·诺特把这些成果综合起来，发明了现代环论。

马克斯·诺特是德国南方城市埃尔朗根（在纽伦堡北边）的一名数学教授。1882年爱米在那里出生。我们可以在日尔曼帝国的发展中了解在其中成长的她的人生，当时是俾斯麦（到1890年任第一任部长和第一首相）和威廉二世（1888年至1918年的德国皇帝）的帝国。威廉统治下的德国是一个异常厌恶女人的男人社会，甚至按19世纪末的标准来看也是如此。德国人形容女性惯用的措辞是Kinder、Kirche、Küche（孩子、教堂、厨房），我想，那些不说德语的人也可以大概推测出女人在社会中的地位。据说威廉二世愚蠢的老婆奥古斯塔·维多利亚皇后认为女人就该这样，只是估计她嘴上不停地唠叨着的是皇帝、孩子、教堂、厨房。要深入了解这一话题，我推荐特奥多尔·冯塔讷1895年的小说《血泪的控诉》（*Effi Briest*）。有文化的人肯定都熟悉伟大的法国作家福楼拜的小说《包法利夫人》（1856年）和俄国作家托尔斯泰的小说《安娜·卡列尼娜》（1877年）中对因生活痛苦不堪而做出有伤风化事情的19世纪女性的描写，但是很少有人知道在这个领域的这部德国作品，即冯塔讷这部朴素的杰作。

因此，当爱米·诺特在18岁左右决定毕生从事纯数学研究时，挡在她面前的是一座险峻的大山。即使她有一个数学家父亲，一个著名大学

教授父亲，情况也是如此。1900年，诺特做出了决定，当时只允许女性作为听众坐在大学教室里，而且还要得到教授的许可。于是1900年到1902年，爱米·诺特坐在埃尔朗根大学的数学教室里，1903年到1904年则在哥廷根大学求学。

到了1907年，出现了一些谨慎的改革，因此诺特获得了埃尔朗根大学的博士学位，这是德国大学授予女性的第二个数学博士学位。这第二个博士学位的获得本应使她获得在大学任教的资格，然而，这种“资格”还没有向女性敞开大门。她在埃尔朗根工作了八年，一直是没有薪水的博士生管理人和临时讲师。但没有什么能够阻止她发表文章。因为在数学方面的出色表现，她很快就成了名人。

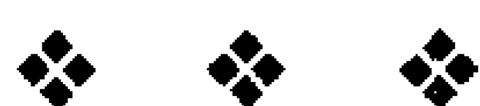
这是爱因斯坦于1905年发表他狭义相对论后几年间的事情。当时爱因斯坦正全力投身于他的广义相对论的研究中，目标是把引力纳入到他的讨论范畴。然而，还有一些困难没解决。1915年6月和7月，爱因斯坦在哥廷根大学的某些讲座中提出了他的广义相对论和一些没有解决的问题。爱因斯坦是这样评价这次活动的：“令我感到高兴的是，我成功地说服了希尔伯特和克莱因。”

这的确值得高兴。即使当时的希尔伯特和克莱因已经过了职业颠峰期（希尔伯特53岁，克莱因66岁），但他们依然是数学巨人，而爱因斯坦则不过是一个有些许成就的少壮派而已，当时他才36岁。当然，希尔伯特和克莱因是在爱因斯坦于1915年来讲座之前就很有兴致地关注了爱因斯坦思想的发展。在爱因斯坦的“劝说”下（大概主要是相信了爱因斯坦是在正确的路线上），他们对广义相对论中的突出问题投入了关注。他们知道爱米·诺特在相关领域的一些工作，因此邀请她到哥廷根来。

[那些涉及变换中不变量问题的相关领域，我将在下面讲述这些思想。相对论中的重要变换就是洛伦茨变换，这一变换告诉我们，当从一个参照系移到另一个参照系时，坐标（即三个空间坐标和一个时间坐标）是如何改变的。在这一变换之下的不变量是“原时”，$x^2+y^2+z^2-c^2t^2$，至

少在进行微积分计算所需的无穷小层次上是这样的。]

诺特按时到达了哥廷根。几个月后，她就发表了一篇论文，解决了广义相对论中的一个未解问题，并给出直到今天物理学家仍然珍爱的一个定理。爱因斯坦自己十分赞赏这篇论文。爱米·诺特成功了。



此时，人们已经承认爱米·诺特是一名一流数学家，但是她的职业困扰依然没有结束。第一世界大战开始的第二年，爱米·诺特的弟弟弗里茨（也是一名数学家）参军。威廉时代各大学在某种程度上有一定的自由，但是哥廷根仍拒绝给予女性以权力。希尔伯特是一个开明的人，他判断一名数学家只凭其才能，因此他勇敢地为诺特争取权力，但是没有成功。

交战双方的某些争论已经成为数学家之间的传奇故事。权力方："当我们的战士返回学校发现他们拜在女人之下学习时，他们会怎样想呢？"希尔伯特："我不明白为什么候选人的性别竟成了反对她作为无薪讲师（即教师的薪水直接由学生的学费支付）的论据。毕竟我们这是大学，而不是浴场。"[9]

希尔伯特对诺特问题的解决方案很独特：他以自己的名字开设课程，然后让诺特去教这些课程。

终于，在第一次世界大战战败后德国社会的全面自由化环境下，对于一位女性来说，"获得资格"并得到大学教师的职位已成为可能，只是作为无薪教师，要依靠学生的学费支付课时费。1919年诺特获得了教师资格。1922年她在哥廷根获得了带薪职位，但是她没有任期，并且少得可怜的薪水很快就不足以应付通货膨胀。

就是在战后最初的几年间，诺特整理了关于环的所有工作成果，把它们转化成为清晰的环论。她于1921年发表的论文*Idealtheorie in Ringbereichen*（"环–域中的理想理论"，当时还没有确定这些术语）被认为是现代代数史上的里程碑，它不仅给出了交换环[10]内部结构的关键结

果，还展示了很快被其他代数学家接纳的这一课题的研究方法，即成为现代代数的严格公理化方法。

范德瓦尔登说："在哥廷根，对我来说最重要的事情是我与爱米·诺特成为熟人，她已经给代数换了新装，比以往任何研究都更一般化……"

到了19世纪30年代初，爱米·诺特已经成为哥廷根研究人员的中心人物。但是她的职位仍旧很低，薪水可怜，而且没有终生职位，但是她是一位权威的数学家这一点不容置疑。以那个时代的观点来看，诺特不是一位标准的女性，她的同事这样说是比较公正的，即使现代人看来也是如此。她身材矮小，略胖，相貌平平，总是带着厚厚的眼镜，声音低沉刺耳。她穿着不合体的衣服，剪成小平头。人们常说她的讲课风格令人难以理解。她的同事都是男性，他们对她既敬畏又喜爱。威廉皇帝时期才过去十几年，那时的人们有自己的友爱表达方式，也许今天的人无法接受。

因此，贬低妙语不见得是不友好，所有对诺特的贬低妙语都变成了数学传说的一部分。比如众所周知的就是，当有人问爱米·诺特的同事兰道他是否认同诺特是一位伟大的女性数学家时，兰道的回答是："爱米肯定是一位伟大的数学家，但要说她是一个女人，我不能肯定。" 诺伯特·维纳对她的描述更诙谐一些："精力充沛近视眼的洗衣女，她的很多学生围拢在她的周围，就像一群小鸭子围拢在一只仁慈的老母鸡的周围。" 赫尔曼·外尔更绅士地表达了大众的观点："优雅从未主宰过她儿时鼾睡的摇篮。" 外尔还试图这样解释人们对她的称呼"Der诺特"（在德语中der一词是定冠词的男性格式）："如果我们在哥廷根……称她是Der诺特，那一定是表示我们由衷地承认她是一名强大且富有创造力的思想家，她几乎打破了性别的障碍……她是一名伟大的数学家，最伟大的数学家。"

即便诺特在哥廷根薪水可怜且没有保障，1933年当纳粹上台时她还是丢掉了这份工作。之前因为是女性而被阻止到大学任教，而此时则因

为是一个犹太人而受到更多的限制。以希尔伯特为首的她的非犹太人同事和旧同事多次为她请命，也都不受重视。

在纳粹统治期间，犹太科学家和反纳粹的天才一般有两种逃脱方法：去苏联或者去美国。爱米的弟弟弗里茨选择了前者，在西伯利亚的一所学院得到一份工作。爱米则走上了后一条路，在宾夕法尼亚州的布林茅尔学院获得一个职位。她的英语还凑合，而且只有51岁，这所学院非常高兴得到这样一位数学天才。然而仅两年后，爱米·诺特因为子宫肿瘤切除手术引起的栓塞而去世。爱因斯坦为《纽约时报》写了讣告，其中的一段如下。

> 在代数领域……那些最有天赋的数学家已经研究了几个世纪的这一领域，她发现了非常重要的方法……世上只有少数人，他们在生命之初就认识到开启人类最美妙、最令人满意的经历不是来自外部，而是与个人的情感、思想和行动的发展密切相关。真正的艺术家、研究家和思想家就是这种人。他们默默无闻的私人生活虽然与常人一样，但是他们努力的成果却是这一代人给他们的后辈创造的最有价值的贡献。

注　解

[1] 大约在1585年，丹麦代数学家西蒙·斯蒂文第一次注意到了整数与多项式之间的这种关系，或者至少第一次对此加以论述。对不起，正文中我没有单独介绍斯蒂文，他是伟大的十进制传播者，为了使它们被欧洲人接受做了大量的工作。他写了一本这方面的书，正是这本书激励了托马斯·杰弗逊，也才有了十进制在新生美国的流行，我们有了“十分钱”（dime）这个名词也应该间接地归功于他。

[2] 职业代数学家如果看到我这样为环下定义，一定会怒不可遏。是的，我陈述得过于简单了，尽管只是过了那么一点点。事实上，环的代数记法要比我上面所给的例子所蕴含的意义宽泛得多。例如，一个环不必有乘法恒等元，即1，而整数 $\mathbb{Z}$ 和多项式都有1。尽管加法是可交换的，但乘法不必是可换的。然而本

书不是一本教科书，我只是想讲清楚大致的意思。

[3] 用这个例子阐述环论不太直观，注意在形如$a+b\sqrt{5}$的数组成的环中，其中a和b都是普通的整数，$9+4\sqrt{5}$是这个环的单位元，即它可以整除1。读者可自行验证。

[4] 定义正则素数没有简单的方法。一个稍稍简单的定义方法如下。如果它不能整除伯努利数$B_{10}, B_{12}, B_{14}, B_{16}, \cdots, B_{p-3}$的分子，那么素数$p$就是正则的。（我已经在第9章介绍了伯努利数。）例如，19是正则素数吗？只要它不能整除伯努利数B_{10}、B_{12}、B_{14}、B_{16}的分子，即5、691、7和3617，它就是。19的确不能整除它们中的任何一个，所以19是正则素数。第一个非正则素数是37，它可以整除B_{32}的分子，这个分子是7 709 321 041 217。

[5] 即如今波兰西部的扎里。布雷斯劳是现在波兰的劳克斯劳。第二次世界大战后，整个德国—波兰边界向西移动了。

[6] 柏林这个数学中心迷漫着古老的日尔曼人的浪漫主义气息。克罗内克曾说："我们是诗。"魏尔斯特拉斯认为"不懂诗的数学家永远也不会成为真正的数学家"。

[7] 证明：$(15m+22n)$取遍所有整数这个假设不是真的。假设存在某个整数k，对于任意m和n它都不等于$(15m+22n)$。重新把$(15m+22n)$写成$[15m+(15+7)n]$，这也等于$[15(m+n)+7n]$。k不能用这其中任意一个表示，即15的倍数加上7的倍数。但是观察一下我的做法：我用一对新数（比原来的数对小且不等于原来的数对：15和7）取代了原来的一对数（15和22）。显然我可以用这种"减少的方法"一直做下去，直到无法做下去为止。欧几里得就是这样证明初等算术的，所以如果我这样做，最后我将得到一对数$(d,0)$，其中d是我原来的一对数的最大公因数。15和22的最大公因数是1，所以我的讨论到此结束：k不能表示成$(1\times m+0\times n)$，对于任意的m和n都成立。这是荒谬的：$k=1\times k+0\times 0$。我用反证法证明了我的结论。

[8] 我参考的是汉纳克所著的《伊曼纽尔·拉斯克传记：一个国际象棋大师的生活》（1959年出版），据说这是最权威的拉斯克传记。我手中的版本是1959年海因里希·弗兰克尔的翻译本，其中有爱因斯坦所写的前言。这真令人羡慕，就如同达芬奇为你的书做插图一样。

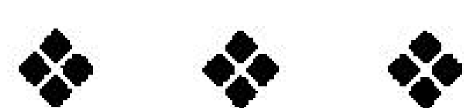
[9] 希尔伯特思想开明，为人热情，怎能不让人崇敬？康斯坦丝·瑞德在1970年为他写了英语传记。

[10] 对于现代代数学家，"交换"和"非交换"表明两种不同的代数品味，对应两种不同的代数应用。我要讲的是代数，所以就不介绍这种品味上的差异了，我没有详述交换和非交换的差别，而选择略过这些基础概念。

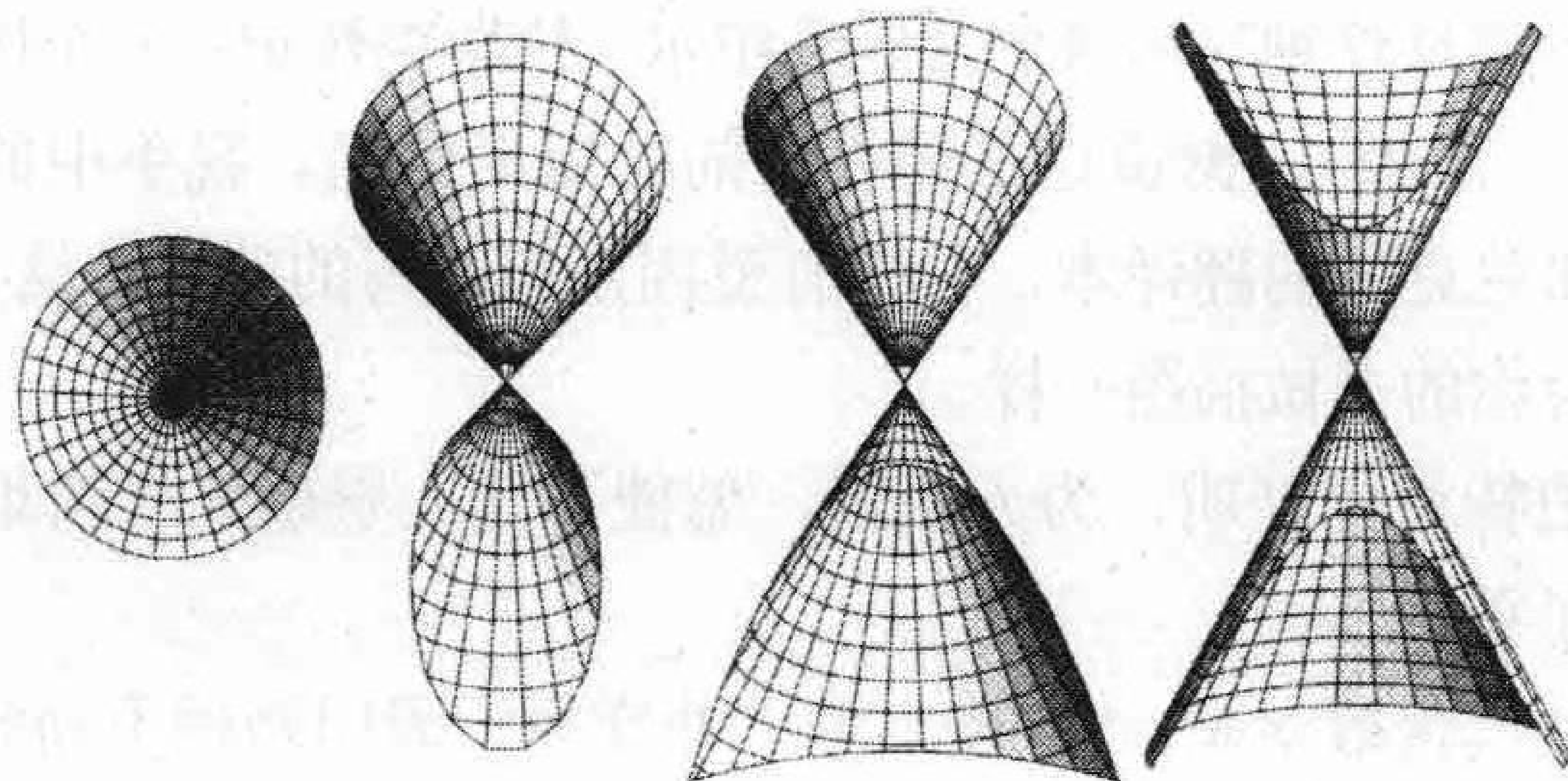
数学知识

代 数 几 何

我将在下一章介绍几何对现代代数产生的极其重要的影响，并将尝试揭示这一影响的实质和进展。在这里我只想介绍几个关于代数几何的基本思想。我将遵照悠久的数学传统，以圆锥截面为例展开本话题。



圆锥截面通常只称为“圆锥”，是一组平面曲线，用平面横切一个圆锥体时就能得到圆锥截面。（还要注意，数学上认为圆锥体不停留在它的顶点，而是朝两个方向上无限延长。）在图AG-1中，这个横切的平面是你阅读的纸面，你必须想象它是透明的。这个圆锥体的顶在这个平面的后面。在第一个图中，圆锥的轴正好与这张纸成90°，所以截面是一个圆。然后，我旋转这个圆锥，把圆锥的远端向上提，于是这个圆变成椭圆。当我再进一步向上旋转这个圆锥时，这个椭圆的一端“走向无穷”，因此得到的是一个抛物线。旋转超过那个顶点时，得到两部分曲线，称为双曲线[1]。



图AG-1　横切这张纸的一个圆锥形成的圆锥截面

当然，这些都是几何。为了得到代数，我们必须使用笛卡儿坐标。对定义圆锥的无穷多个点，笛卡儿的x、y坐标满足一个关于x和y的二次方程，这个方程是

$$ax^2+2hxy+by^2+2gx+2fy+c=0$$

（上面方程的系数a、h、b、g、f所用字母的选择可能看起来有点奇怪，稍后我会解释。）表述这同一件事的另一种方法是：一个圆锥就是某两个未知量的二次方程的零点集合。它是满足以下条件的点(x,y)的集合：它使得方程的计算结果等于零。

例如，图AG-2a给出的椭圆有下面的方程：

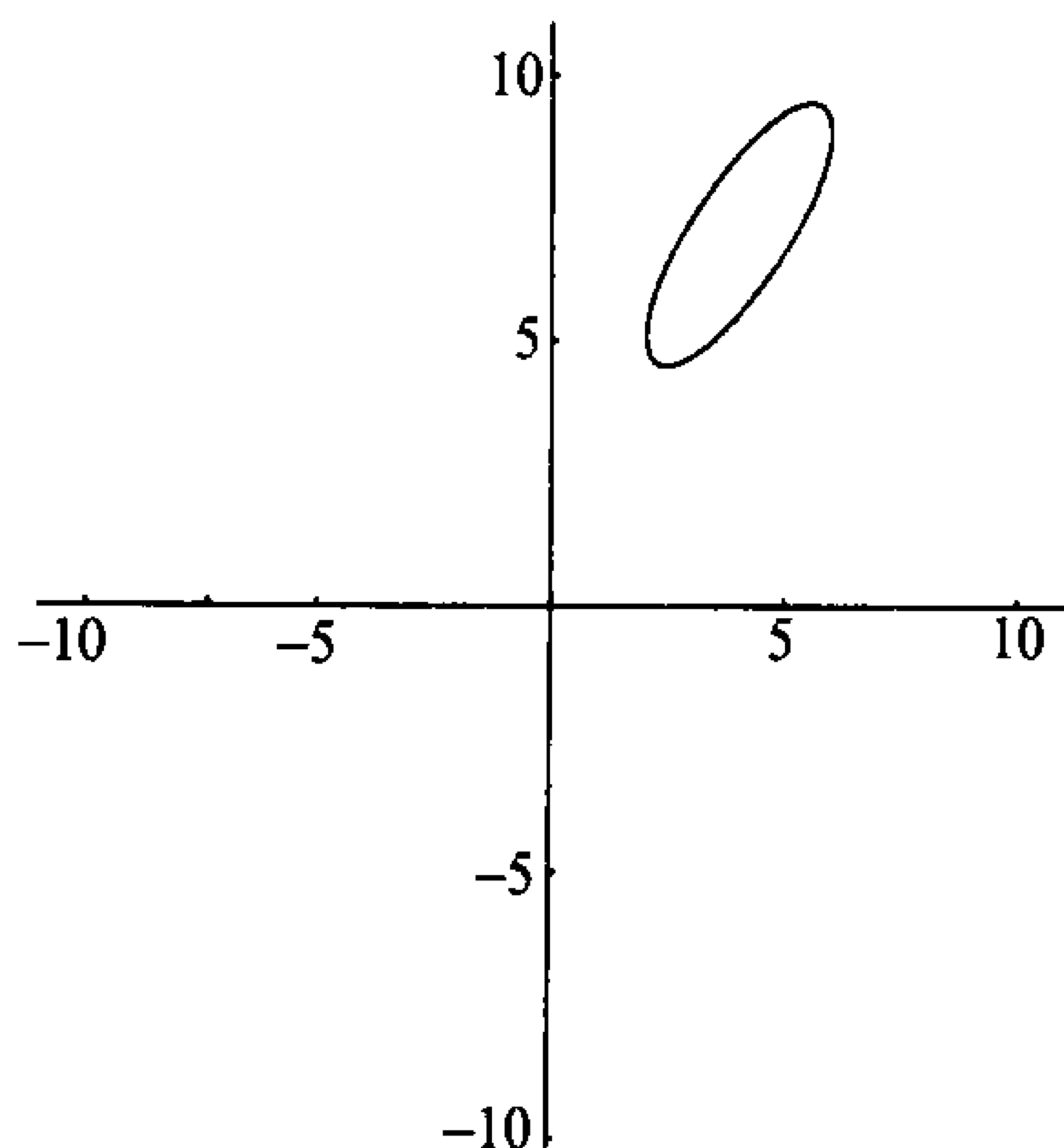
$$153x^2-192xy+97y^2+120x-590y+1\,600=0$$

现在，假设我要把这个椭圆移动到这个平面的其他地方，然后再如图AG-2b那样稍微旋转一下，那么这个代数方程会发生什么样的变化呢？

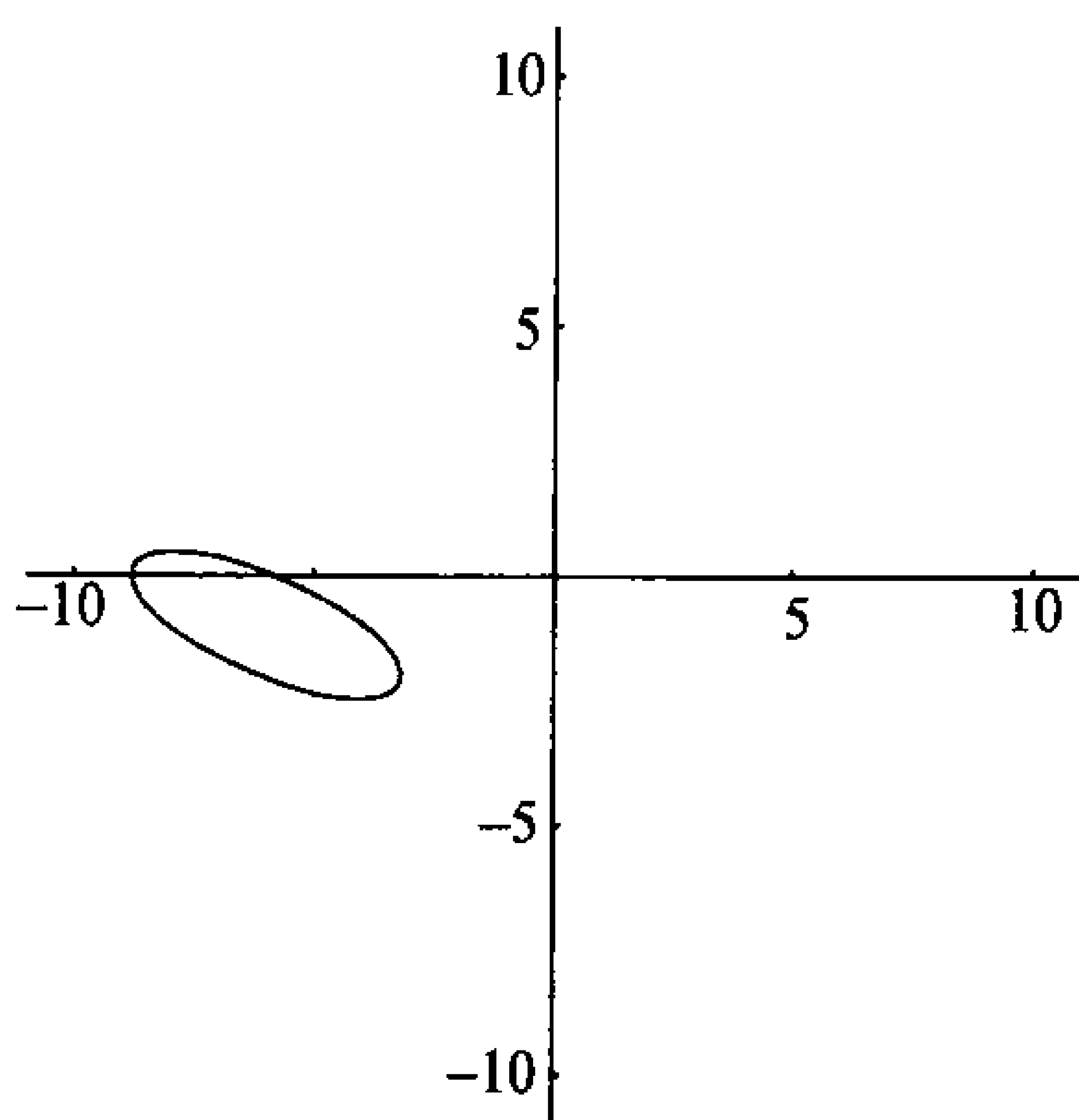
当然这个方程是会发生变化的。对于这个椭圆，它的新方程是：

$$369x^2+960xy+1321y^2+5\,388x+8\,402y+18\,844=0$$

然而，它仍然代表同一个圆锥。它的大小和形状没有变。它没有变大也没有变小，没有变胖也没有变瘦。



图AG-2a



图AG-2b

图AG-2　椭圆示例

于是就产生了下面这样有趣的问题：在这两个代数表达式中，什么告诉我们它们给出的是同一个圆锥曲线呢？从一个方程到另一个方程的变化过程中，什么没有改变，或按数学所说的，什么是不变量？

答案不是一眼能看得出来的。所有系数的大小都改变了，而且符号也发生了变化（负变正）。但是，不变量还是隐藏在那里。再次给出方程的一般形式：

$$ax^2+2hxy+by^2+2gx+2fy+c=0$$

计算下面这些量：

$$C=ab-h^2$$

$$\Delta=abc+2fgh-af^2-bg^2-ch^2$$

在我的两个椭圆方程中，上面的计算结果是C分别为5 625和257 049，Δ分别为−1 265 625和−390 971 529。显然这些也不是不变量。但是仔细观察一下：计算两种情况下的Δ^2/C^3，答案都是19！无论把这个椭圆移动到这个平面的什么地方，如何旋转它，无论由此而得到什么样的方程，方程系数计算的结果都是19。这个数值Δ^2/C^3是一个不变量。事实上，如果取这个数的平方根然后再乘以π，我们就得到了这个椭圆的面积，在平面上移动这个椭圆时，它的面积显然是不变的。

$$椭圆的面积=\pi\times\sqrt{\Delta^2/C^3}$$

下面是另一个不变量。使用系数a、b和我刚才计算的C求下面二次方程的两个根：

$$t^2-(a+b)t+C=0$$

用较大根除以较小的根。再用1减去这个结果。再取平方根。通常称这个数为e（或者ε，以便与欧拉常数e=2.718 281 828 459…相区分，欧拉常数是自然对数的底），它衡量的是这个椭圆的偏心率，即这个椭圆偏离圆的程度。如果e=0，那么这个椭圆就是一个圆；如果e=1，这个椭圆实际上是抛物线。显然这应该是一个不变量，它确实是。如果计算我上面描述的椭圆的两个方程的e，结果e（或ε）是$2\sqrt{2}/3$，大约是0.942 809 04，

与圆相比更靠近抛物线。可以想象，这个椭圆是一个长长瘦瘦的椭圆。

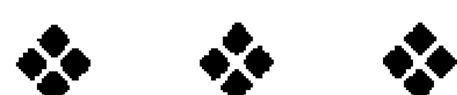
这个椭圆的实际维度是多少？当移动椭圆时，维度不发生变化。它们也是跟系数无关的不变量吗？是的。对应于刚才我计算的不变量 Δ^2/C^3，这个椭圆的长轴是：

$$2\times\sqrt{\sqrt{I\div\left(1-e^2\right)}}$$

短轴是：

$$2\times\sqrt{\sqrt{I\times\left(1-e^2\right)}}$$

如果计算我的两个椭圆方程的轴，那么两个方程的长轴都是6，短轴都是2。同猜想的一样，这些数是不变量。



关于笛卡儿几何我已经介绍了很多，因为它不仅让我们看到了不变量的思想，同时也让我们看到了现代代数中其他重要的思想。

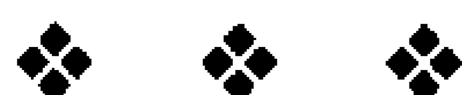
在圆锥曲线讨论中还隐含着矩阵的思想。对于我前面给出的圆锥曲线的一般方程，这个重要的矩阵是：

$$\begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix}$$

学数学的学生通常用这种方法记住这个矩阵：All hairy guys have big feet (guys也作girls，看你的喜好)。[①]我们能够像在第9章所描述的那样求出这个矩阵或任何正方形矩阵的行列式。这个矩阵的行列式就是我上面定义的 Δ。

如果给定一个圆锥曲线在某个笛卡儿坐标系下的方程，它的 Δ 的值等于0，那么该圆锥曲线是一个“退化的”圆锥曲线：它或者是一对直线，或者是一条直线，或者是一个点。可以验证对于点(0,0)和方程 $x^2+y^2=0$，Δ 的确等于0。

① 看这句话的首字母。——译者注



为了看一下另外一个论题，我要重复前面我提到的问题：为什么圆锥曲线的一般方程通常有相当奇怪的系数a、h、b、g、f、c。没有d和e的原因容易解释：d可能与它在微积分中的使用dy/dx相混，而e可能与欧拉常数e混淆。但是为什么所有字母的顺序是这样的呢？为什么不写成下面这种形式的一般方程呢？

$$ax^2+bxy+cy^2+fx+gy+h=0$$

答案很简单。笛卡儿的x和y几何的确不是研究圆锥曲线的最好工具。事实上，如果允许存在无穷远点，那么圆锥曲线更适合用代数处理，笛卡儿几何是不合适的。我给出的圆锥曲线的一般方程已经超出了那种经典几何，它允许存在无穷远点。

对非数学人士来说，术语“无穷远点”[2]听起来似乎有点令人困惑。其实这只是引入到几何中用来简化事物的一种艺术表达而已。例如，如果允许无穷远点，那么平行线难以处理的地方就消失了。把原来的定义：

> 平面内任意两条直线都将相遇在一个点，除非它们平行；平行时它们将永不相遇。

更改为：

> 平面内任意两条直线都将相遇在一个点。

对此的理解是可以把平行线考虑成为相遇在一个无穷远点。你也许不明白这种做法的重要性，但是不可否认它变得简单了。

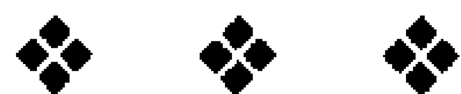
但是，在平坦的欧几里得平面内，古老的笛卡儿坐标系不能处理这个问题。如果用笛卡儿坐标写出一个无穷远点，应该写成(∞,∞)。这只是一个点。人们的直观感觉是，如果一对平行线在一个无穷远点相遇，那么与它们不平行的另外一对平行线应该相遇在一个不同的无穷远点。换句话说，我们需要很多无穷远点。

表示无穷远点的另一种作法就是用三个坐标取代二个坐标。不用坐标(x,y)确定平面上的点，而是用坐标(x,y,z)确定平面上的点。在这里我们要采取一些措施，以防止这种表示法与真正的三维几何冲突，所以下面是一个限制条件：我们认为x、y、z有相同比例的所有点代表同一个点。例如点(7,2,5)、(14,4,10)和(84,24,60)都代表同一个点。这不是什么新想法，小学的时候我们就已经知道3/4、6/8、9/12、30/40等表示相同的分数。这一限制是把维度压回到二维。

另一种考虑这种新坐标的方法是：用x/z和y/z取代x和y。如果z是0，x/z和y/z当然无法计算：它们就是无穷远。新的三坐标系避免了这些小麻烦。我们确定无穷远处的点，只需通过让z是零来表明那些点。此时，我们可以认为$(x,y,0)$、$(2x,2y,0)$、$(3x,3y,0)$以及所有其他$(kx,ky,0)$都是相同的无穷远点，只不过标签不同而已，而且有很多这样的点，不只一个。

事实上，它们都在一条直线上，这是方程为$z=0$的直线。它被命名为“无穷远线”，再遇到这样的术语你应该不会惊讶了。只有唯一一条无穷远线，它是由很多无穷远点组成的，每一个无穷远点都有不同的坐标。

这种在普通的笛卡儿几何中加入了一条无穷远线的新几何称为射影几何。我现在描述的新坐标系是让射影几何满足某种算法。但最纯粹的射影几何与坐标无关。它只与那些被投影后仍然为真的几何原理有关。想象一个幻灯片上的几何图形。这个幻灯片与平坦表面成一定角度，来自点光源的光照射它，于是这个几何图形就被投影到这个平坦表面上。在此过程中，一些几何性质丢失了。例如，圆不再是圆，它变成了圆锥。然而，有些性质还是保留了下来。我将在下一章详细讨论这一问题。



在这一新坐标系下，圆锥曲线的一般方程变成什么样子了呢？让我们尝试在原来的圆锥曲线的一般方程中分别用x/z和y/z取代x和y：

$$a\left(\frac{x}{z}\right)^2+2h\left(\frac{x}{z}\right)\left(\frac{y}{z}\right)+b\left(\frac{y}{z}\right)^2+2g\left(\frac{x}{z}\right)+2f\left(\frac{y}{z}\right)+c=0$$

在上面这个方程的两边乘以z^2，就得到：

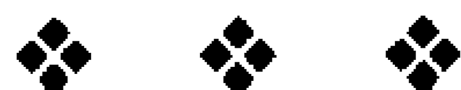
$$ax^2+2hxy+by^2+2gzx+2fyz+cz^2=0$$

稍微改变一下各项的排列顺序：

$$ax^2+by^2+cz^2+2fyz+2gzx+2hxy=0$$

此时，使用a、b、c、f、g、h顺序的理由就很明了了。在这个新坐标系下，我们有x^2、y^2、z^2项，有yz项（即省去了x的xyz项），有zx项（即省去了y的xyz项），有xy项（即省去了z的xyz项）。请看，这是对称的！[3]

从严格的数学观点看，我们在这里得到的不是真正的对称，只是齐次的。这种坐标称为齐次坐标。尽管如此，这毕竟是在正确的方向上迈进了一步，同时也展示了对称的概念在现代数学中有多么大的吸引力。



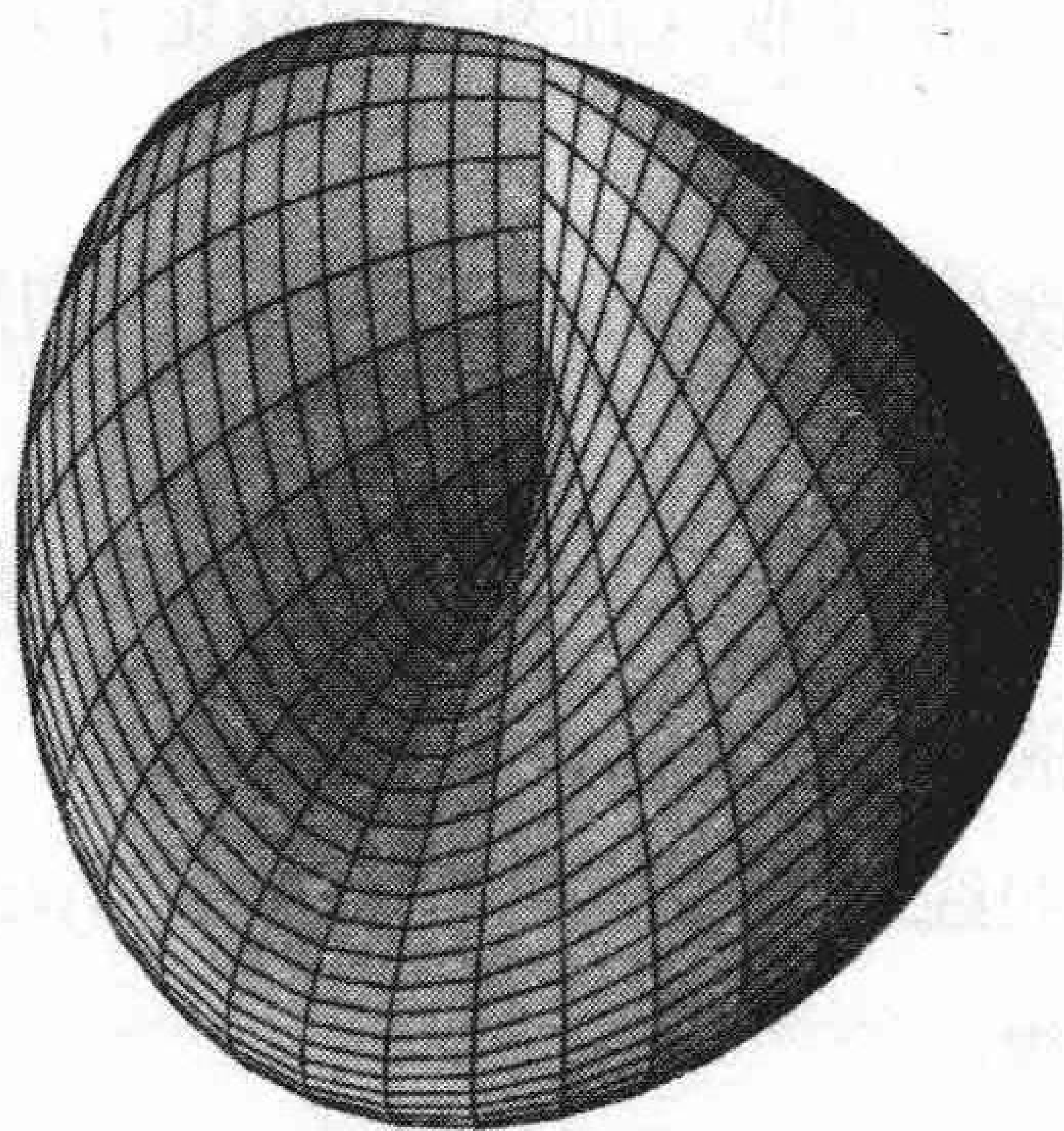
新坐标系引出了现代数学中另外一个重要的话题。一旦开始研究这种新的布局，你会发现它很巧妙，不可思议，不同于我们熟悉的平坦的欧几里得平面。所谓新布局就是当我们加入了无穷远点和无穷远线时得到的布局。例如，无穷远线的另一端是什么呢？再问另外一个问题：给定一对平行线，我们已经声明它们相遇在无穷远处的一个点，如果想要到达那个点，应该沿着这些线在哪个方向上前进呢？如果这一对平行线是东西向，那么这个无穷远处的点离东远还是离西远呢？

这一类问题表面上很幼稚，像是小孩子的问题，但是事实上它们非常重要。实际上，这些问题把我们引入了拓扑的领域。

在普及性的数学课本中，拓扑一般是作为“橡胶几何”给予介绍的。拓扑学家对图形的这样一些性质感兴趣：当这些图形在任意方向以任意大的力量被拉伸，只要不被撕裂、切割时，保持不变的性质。例如在这些规则下，球的表面等于立方体的表面，但是它不等于甜甜圈的表面。而甜甜圈的表面等于带把咖啡杯的表面。

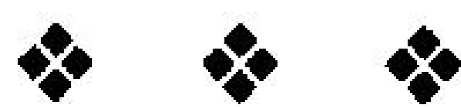
从拓扑学上讲，古老的欧几里得平面等于缺少了一个点的球平面。（想象一下把这个平面“折叠”成端坐其上的一个球，但是这种折叠不会

在北极真正相遇。）然而，加上那个点构造一个完整的球不会得到新布局[4]。那个缺失的点对应于无穷远点。而我们的新布局，即数学术语中的投影平面，没有这样的唯一点，球面有无限多个这样的点。所以投影平面在拓扑意义下不等于完整球面。在拓扑的意义下它等于一个更特殊的对象：一个有折缝的球（图AG-3）。



图AG-3　拓扑意义下的投影平面

注意，这个对象与莫比乌斯带一样只有一个面。如果让一只蚂蚁在其上爬行，让它一直向前爬穿过这些折痕，那么它能访问这个表面的每一个点，无论是里面还是外面。看，一个如此幼稚的问题就把你引到了投影平面的领域！

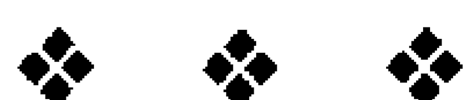


不变量、矩阵及其行列式、对称、拓扑，这些都是现代代数的关键概念。事实上，我还没有完成用代数方法对圆锥的处理。

我说过，在这个平面上移动这些椭圆。这就是我观察我在做什么的一种方法。还有一种方法，就是认为椭圆静静、稳固地坐在它的平面里，而坐标系是移动的。（你可以这样想，在一张透明胶片上打印x轴、y轴以及方格坐标，它在基础平面上滑动。）

这两种方法就是变换的例子，这是现代数学中另外一个非常重要的

概念。这些特殊的变换，即只改变位置和方向，而不改变距离和形状的变换，称为等距变换。它们本身就构成了一个研究领域，我将在第13章中详述。还有更复杂的变换：仿射变换（允许把某些直线拉伸和“剪切”，可以把矩形变成平行四边形）、投影变换（像上文所说的那样对图形做投影）、拓扑变换（随心所欲地拉伸或挤压，但不能剪切）、洛伦兹变换（在狭义相对论中）、莫比乌斯变换（在复变函数论中）以及众多其他变换。



对于我提到的齐次坐标再多说一点。这里我们用三个数(x, y, z)来确定二维空间中的一个点。

在高中代数中，我们知道直线的笛卡儿坐标方程是$lx+my+n=0$。它的齐坐标方程是什么样的呢？

如我在介绍圆锥曲线时所做的那样，用x/z和y/z分别代替x和y，然后化简，得到下面的方程：

$$lx + my + nz = 0$$

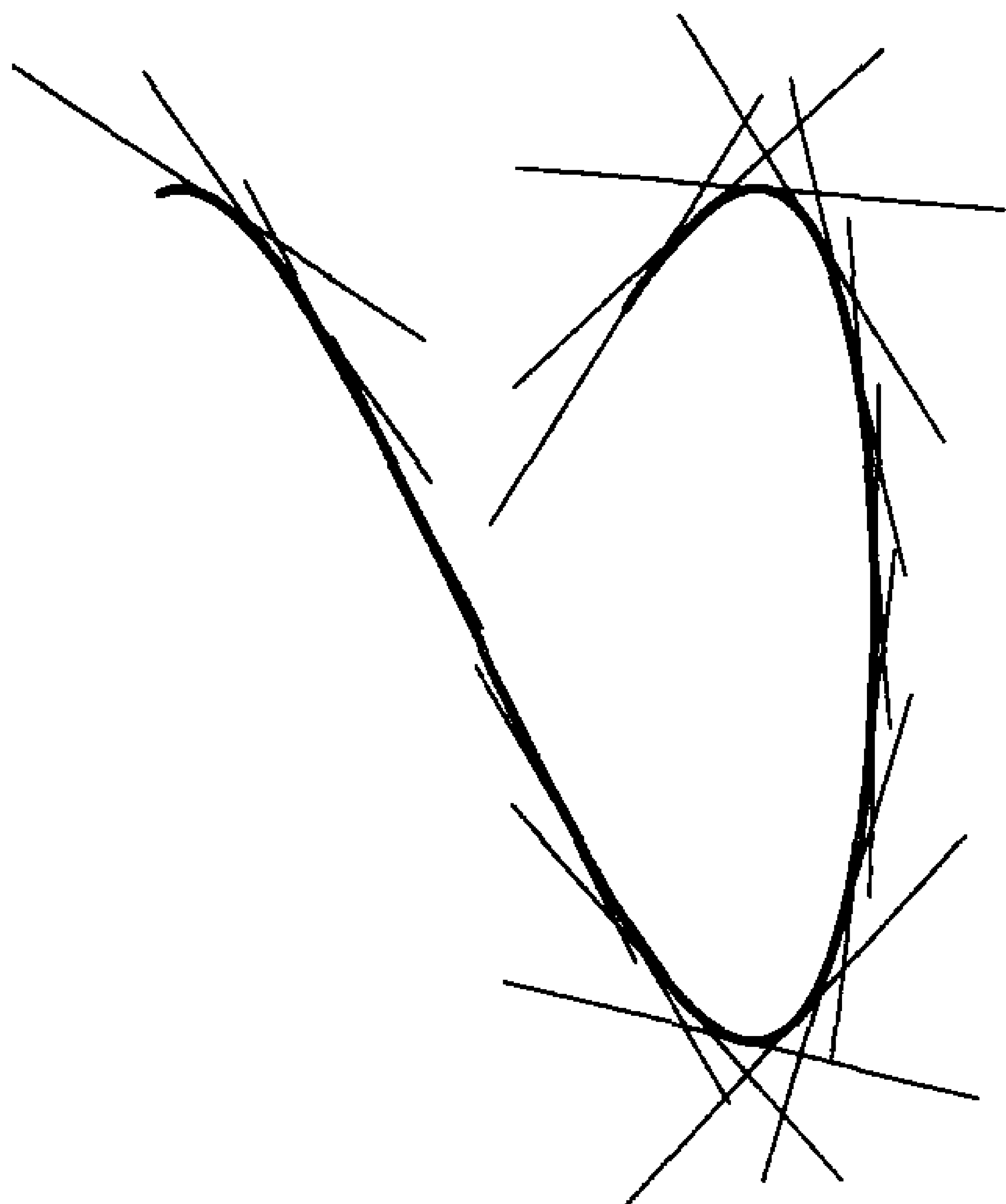
这就是直线的齐次坐标方程。但是观察一下：这个方程表明直线是由三个系数(l, m, n)决定的，就像一个点是由它们三个齐次坐标决定的一样。多么对称啊！

于是引发了这样的问题：在齐次坐标系下，我们能否围绕着线而不是点来构建几何呢？毕竟，就如线是由无穷多个满足某个线性方程$lx+my+nz=0$（其中l、m、n是固定的）的点构成的一样，点是无穷多条经过这个点的线构成的！其中每一条线都满足方程$lx+my+nz=0$，但是，现在点坐标x、y、z是固定的，而系数l、m、n却取无穷多个值，每一组值都要画一条经过点(x,y,z)的“铅笔线”。

类似地，我们不把类似于圆锥曲线这样的曲线考虑成一个移动着的点的轨迹，而是把它看成一条移动着的线的轨迹，如图AG-4所示。

我们能够根据这一思想构建几何吗？是的，可以。这种“线几何”实际上是由德国数学家朱利叶斯·普吕克于1829年开创的。我们还是接

着叙述历史故事吧。



图AG-4　由线而不是点定义的曲线

注　解

[1]“椭圆”、“抛物线”和“双曲线”等名词都是阿波洛尼乌斯给出的。

[2] 文艺复兴时期的画家们在解决投影问题时，也许想到了无穷远点这一概念，但是大约在1610年才由天文学家开普勒把无穷远点引入到了数学中。开普勒把一条直线想象成一个圆，而这个圆的中心就在无穷远处，这个概念可能来自于他的光学著作，在光学里想象无穷远点就变得很自然了。

[3] 事实上，不是所有作者都沿用这种方法。例如，迈尔斯·里德在他的大作《代数几何初步》（1988年由剑桥大学出版社出版）中把非齐次二次多项式写成$ax^2+bxy+cy^2+dx+ey+f$的形式。

[4] 在欧几里得平面内加入那个缺失的点就会得到一个黎曼球，这对思考复变量函数很有帮助，真应该感谢这个无穷远点。

第13章

几何复苏

任何一所大学数学系的玻璃橱柜里总会摆放着一些数学模型，通常会有几个多面体，有凸起的，有星状的（参见图13-1），也可能有用纤维做的规则表面、用来说明不同球体包装问题而粘合在一起的乒乓球、莫比乌斯带、克莱因瓶和其他各种奇怪的东西。

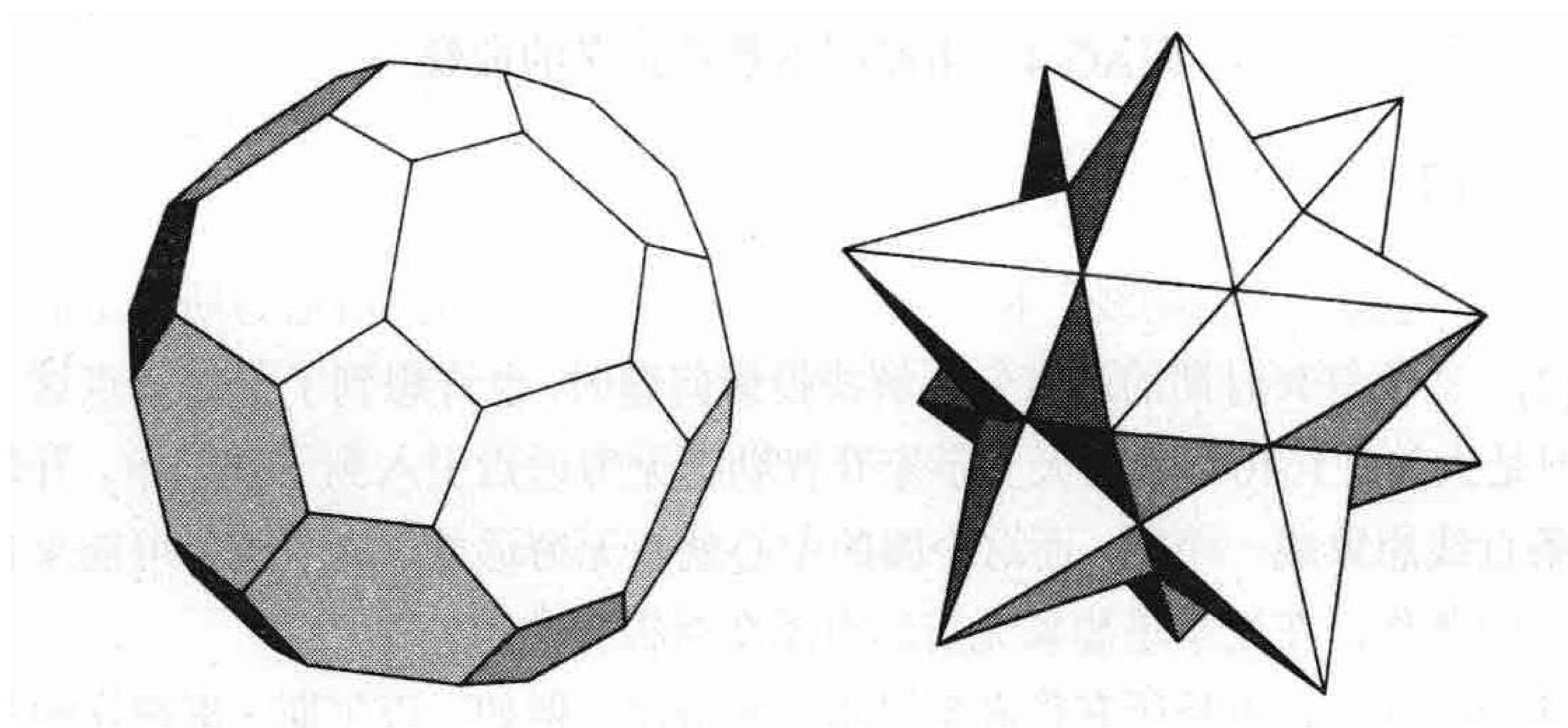
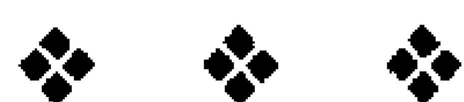


图13-1　凸多面体（左边）和星状多面体（右）

现今这些模型多少失去了吸引力，落满了灰尘。Maple和Mathematica等数学制图软件出现后，用这些软件可以一瞬间在电脑上生成这些模型，而且为了方便观察，可以旋转它们，还可以随意转换、变形，让它们相交，因此再利用木材、卡片、纸和纤维制作这些模型就显得荒唐可笑了。然而在19世纪和20世纪的大部分时间里，对于数学家和学数学的学生来说，制作几何图形的实体模型是一种很惬意的消遣，我很遗憾它没有持

续更长的时间。我自己的青年时代几乎就是在制作马丁·坎迪和罗莱特的经典著作《数学模型》中的所有模型的过程中度过的，那是一段幸福而有意义的时光。我的得意之作就是把五个正方体嵌入到一个十二面体中的卡片模型，每一个立方体涂上不同的颜色。

这种对直观教具和数学思想模型的兴趣是19世纪初等几何复苏的产物。正如我在第10章中提到的那样，尽管在17世纪几何领域取得了一些有意义的进展，但是到了这个世纪的后期，微积分带来的令人兴奋的新思想使得几何变得黯淡无光。到了1800年，几何已不再是有吸引力的数学领域，那个时代的著名数学家很可能只在学校学过欧几里得几何，除此之外没有研究过任何几何。之后的一代人却完全改变了这一切。



19世纪几何学的第一次突破是由法国人让·维克多·庞斯列取得的，他在极其恶劣的条件下取得了这一成就。在庞斯列24岁那年，他作为一名工程兵官员随同拿破仑的军队出征俄国，来到了莫斯科。寒冬的恶劣天气迫使拿破仑军队大撤退。庞斯列没有随大部队撤退，而是被留下来收拾残局，其实就是等死。一个俄国搜索队发现他穿着军官制服，于是把他带离战场审讯。之后，战犯庞斯列不得不开始长达五个月的跋涉，穿过冰冻的河面，前往伏尔加河的萨拉托夫战俘集中营。

在恶劣的关押条件下，为了转移自己的注意力，庞斯列温习其在巴黎综合理工大学所接受的良好数学教育知识，使自己的大脑不得空闲下来。到了1814年9月他可以回到法国时，他已经做了整整七本数学笔记。这些笔记就是他的著作《论图形的射影性质》的初稿，这本著作是现代射影几何的基础教科书。

庞斯列的著作不是纯代数的。事实上，整个19世纪前半叶关于射影几何到底属于分析几何领域还是综合几何领域是有争议的，这场争议现在看起来也挺有意思。笛卡儿遗留下来的分析几何充分利用代数和微积分的力量去揭示几何图形的一些结果，所谓的几何图形是由直线、圆锥

曲线以及更复杂的曲线和表面组成的体系。从希腊人开始经由帕斯卡遗留下来的综合几何比较喜欢纯正的逻辑表述，尽可能不使用数值和代数。

因为它的定理不涉及距离和角度，所以初看起来，射影几何似乎是蹒跚两个世纪的笛卡儿综合方法的复兴。事实上，这只是幻觉。19世纪20年代后，德国数学家莫比乌斯、卡尔·费尔巴哈以及普吕克各自独立地引入了齐次坐标，使得射影几何被完全代数化。

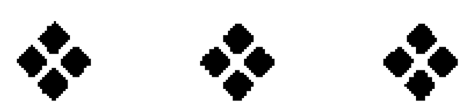
除庞斯列创立了现代几何学基础之外，19世纪20年代还发生了另外一场几何学革命，那就是1928年洛巴切夫斯基在一本俄语地方杂志上发表的一篇论非欧几何的论文。随后他把这篇论文提交给圣彼得堡科学院，但是他们却因为它过于荒谬而拒绝接受。洛巴切夫斯基争辩说，古典几何的共同假设，例如三角形的三个角之和等于180°的假设，也许不是放之四海而皆准的真理，而只是可选的公理。通过选择不同的公理，你也许可以得到不同的几何，这一几何看起来与欧式几何完全不同，它是非欧几何。

年轻的匈牙利数学家鲍耶也正在沿着相同的路线工作。高斯也是如此，1824年他在写给朋友的一封信中提到："三角形三个内角和小于180°的假设带来一个古怪的几何。它完全不同于我们的几何，但又相容……"事实上，高斯几年前就开始思考这些问题。然而，他是一个喜欢平静生活、回避争论的人，所以从来没有发表过他的想法。

这些想法都会引起争议，就如洛巴切夫斯基所经历的那样。欧氏几何的平坦平面（三维空间是"平坦"的空间）中，平行线和平行平面永远不相遇，它的三角形的三个角之和总是等于两个直角之和，它有相似和全等的精妙表示，这些都已深深植根于欧洲人的意识之中，并且得到控制欧洲人思想的伊曼努尔·康德哲学的进一步巩固。在1781年的著作《纯粹理性批判》中，康德指出欧氏几何（用现代术语表示）"深埋于"人类的心灵之中。康德说，我们感知这个世界是欧几里得式的，我们不

能把它感知成别的。按他的意思，就是说这个宇宙是欧几里得式的，欧几里得真理超越了逻辑分析的范畴。[1]

在这样的环境下，19世纪20年代和30年代，数学家们不得不接受奇怪的新几何。这是一场革命，这些数学家被很多康德学派的人认为是危险分子。在那个时代人们对自己信仰的哲学非常认真，当时法国革命和战争的恐怖仍阴魂未散。他们认为哲学的颠覆会带来社会的颠覆。如果庞斯列的射影几何是19世纪几何的第一场革命，那么洛巴切夫斯基和鲍耶的非欧几何就是第二场革命。我们将看到第三场、第四场、第五场革命接踵而至。



在本章前面的数学知识中，我提到了普吕克和他的线几何。普吕克出生于1801年，比阿贝尔大一岁。他的研究生涯很漫长，当过43年（1825—1868）的大学教师，大部分时间都在波恩大学当教授。他的两卷著作《分析几何的发展》（分别发表于1828年和1831年）是那个时代代数几何最前沿的作品，只是其大部分内容使用了老式的非齐次坐标。在19世纪30年代，他开始用另外一种方法研究更高级的平面曲线，这里，“更高级”一词的意思是“次数大于2的代数”，也就是说，代数曲线比圆锥曲线更难处理。

对平面曲线的研究完全是从“分析”的角度来做的，即利用19世纪30年代已经存在的代数的所有资源和微积分来推导制约这些曲线的定律和它们的性质。普吕克1839年的著作《代数曲线理论》对这些曲线的渐近线，也就是这些曲线趋近于无穷时的行为方式，做了权威性的论述。

普吕克的线几何过了很长时间才问世。1847年到1865年这17年间他被聘为波恩大学物理系主任。事实上，对线几何的研究直到他去世的1868年也没有完成，这一研究留给了他的年轻助手克莱因来完成。稍后我再详细介绍这个人。

得助于代数、微积分以及几何的滋养，对曲线的这种兴趣成为19世

纪中期数学的一大增长点。要知道，特别是在数学软件出现之前，在纸上把代数方程转化成曲线是一件非常艰苦的工作，需要进行大量的计算和观察。例如，下面的四次方程是一个相当普通的代数方程：

$$4\left(x^2+y^2-2x\right)^2+\left(x^2-y^2\right)\left(x-1\right)\left(2x-3\right)=0$$

当你标出对应于x的y时，谁会知道得到了如图13-2所示的那样可爱的&形图呢？我是知道的，在我年轻时迷恋马丁·坎迪和罗莱特的经典著作《数学模型》的时候，用铅笔、坐标纸和计算尺我就能画出这条曲线，《数学模型》一书中给出了大量平面曲线和三维空间图形。

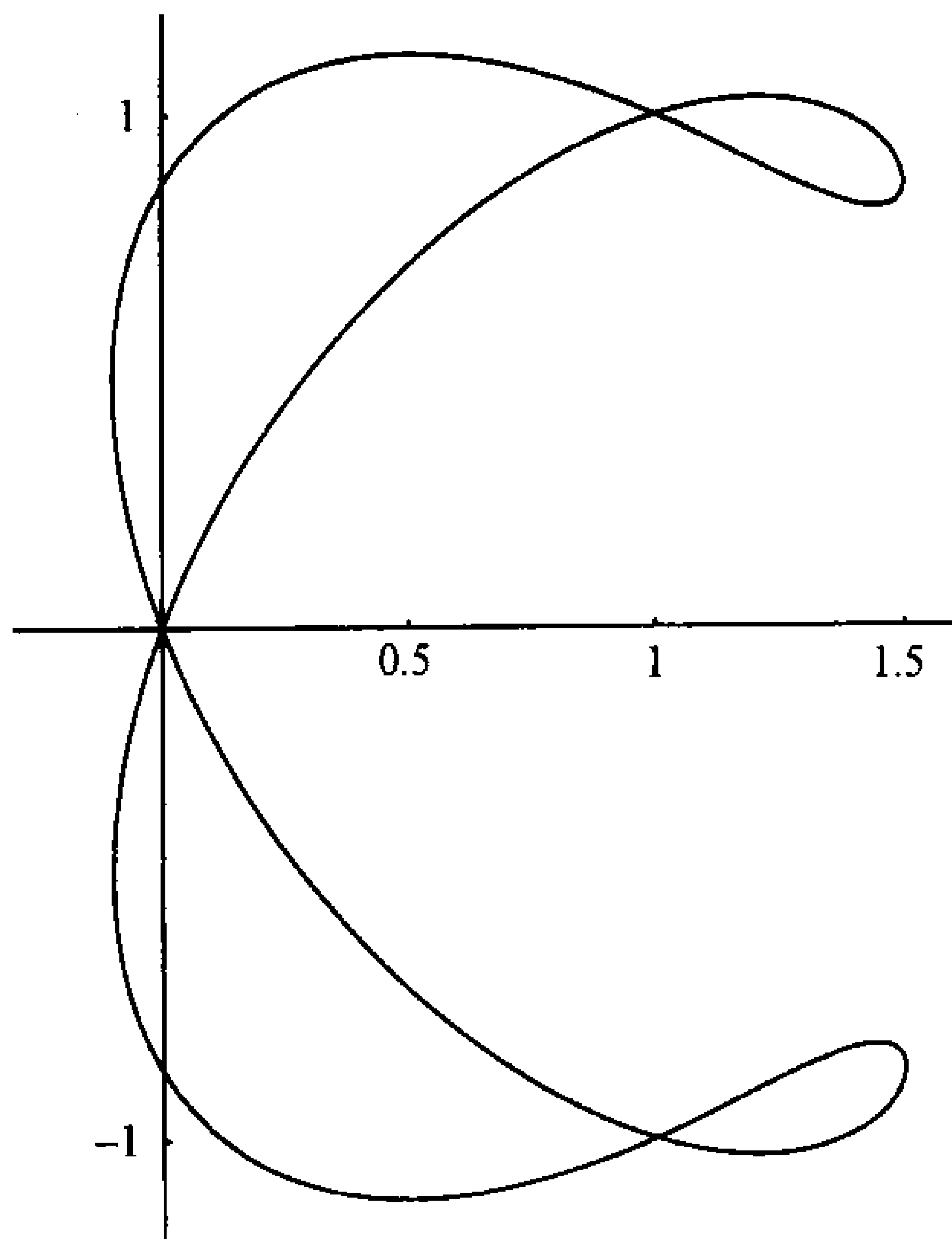


图13-2　&形曲线

此时若有读者怀疑青春期的我不善于社交，应该不算错得离谱。然而，我还是愿意说以我年轻时的偏执，这不复存在的精确的数字计算和绘图实践给我提供了一种特殊的满足感。这不只是我一个人的观点，像高斯这样的人物也有这样的感受。哈罗德·爱德华教授在他的著作《费马大定理》的4.2节中很好地阐述了这一观点，并引用了高斯的话。

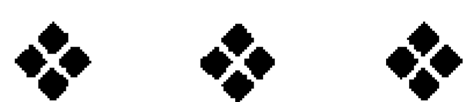
> 如其他伟大的数学家一样，库默尔是“一台贪婪的计算机器”，因此他的发现不是依赖于抽象的思索，而是依赖于特殊计算实例的不断积累。今天，人们对这些计算实践不以为然，计算可以是一种娱乐的想法也很少有人说起。然而，高斯曾经说过，他认为出版二元二次式的分类全表是一件多余的事，“因为，(1)经过少量实践后，任何人都不需要花费太多的时间就能很容易地为自己计算出一个符合特殊要求的表格来；(2)这样的工作本身就具有一定的魅力，所以花一点时间进行计算是一件自娱自乐的事；更重要的是，(3)很少有机会做这样的事。”你还可以举出牛顿和黎曼仅仅为了乐趣而进行长时间计算的例子。任何花时间做（爱德华教授书中的）计算的人都应该会发现，库默尔的计算以及他从这些计算中得到的定理都是他力所能及的，而且他很享受这个过程，只不过没有亲口承认罢了。

顺便说一句，根据《科学传记辞典》的记载，普吕克是一位疯狂的数学模型制作者。

马丁·坎迪和罗莱特不是无中生有地弄出图13-2的。他们是从珀西瓦尔·弗罗斯特的优美小册子《曲线描述》中得到它的。我只知道弗罗斯特出生于1817年，是剑桥国王学院会员，也是皇家学会会员。他的著作首次出版于1827年，此书向读者展示了从一个数学表达式得到一个平面图形的所有可能方法，其中包括一些很巧妙的捷径。注意，弗罗斯特并没有只局限于代数表达式。我自己手中的《曲线描述》是1960年出版的第五版，它的封二上印有正文中所讲述的所有曲线的图片。图13-2的&形曲线是弗罗斯特的图27中的插图VII。

弗罗斯特的著作还回顾了中世纪的代数几何学家，普吕克属于其中最早的一批伟人。与弗罗斯特的《曲线描述》主线相同但数学意义更深的是爱尔兰数学家乔治·萨蒙的四本教科书：《论圆锥截面》（1848年出

版）、《论高阶平面曲线》（1852年出版）、《现代高等代数课程介绍》（1859年出版）和《论三维分析几何》（1862年出版）。在我的桌子上有一本《论高阶平面曲线》。这本书中满篇都是奇妙的术语，但是，令我伤心的是，大部分现在已经不再使用：蔓叶曲线、螺旋线和外旋轮线，蜗牛形曲线和双纽线，甲种尖点和坡式尖点，孤点、歧点和结点，凯莱曲线、海塞曲线和斯坦纳曲线。[2]在萨蒙的时代，齐次坐标已经"安居"下来，但他还是称它们为"三线"坐标[3]。



萨蒙也是一台"贪婪的计算机器"。在他的《高等代数》第二版中，他放入了一个求解六次一般圆锥曲线时得出的不变量。如果你看到了我在数学知识中给出的一般圆锥曲线不变量，就会相信这绝非一般的壮举，它整整占据了他著作的13页。

我提到这些就是提醒人们，中世纪对曲线和表面的这种痴迷得益于纯代数的滋养，反过来它又孕育出了其他结果。我在数学知识中所描述的不变量最初完全是用代数语言构建而成的，几何解释只是衍生的。

凯莱（顺便说一下，他是萨蒙的亲密朋友）和西尔维斯特是1840年以后这一领域的关键人物。他们的大部分工作都是研究多项式不变量，而不是我在数学知识中提到的代表二次曲线的x、y的二次多项式的不变量（如果用齐次坐标，就是关于x、y、z的二次多项式）。所有多项式的集合，比如说未知量x、y、z的多项式，若其系数取自某个域，比如说复数域，那么这个集合在普通的加法和乘法下形成一个环。可以把两个多项式相加、相减或相乘：

$$\begin{aligned}&\left(2x^2-3y^2+z\right)\times\left(y^3z+4xyz^3\right)\\&=2x^2y^3z+8x^3yz^3-3y^5z-12xy^3z^3+y^3z^2+4xyz^4\end{aligned}$$

但不能把它们相除。它是一个环。因此对多项式不变量的研究实际上就是对环结构的研究。

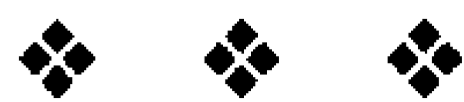
在19世纪，没有人这样想。不变量这条小河开始流入环论这条大

河的最初一幕发生于19世纪80年代末，是保罗·戈尔丹和希尔伯特的功劳。

他们两人都是德国人，但戈尔丹比希尔伯特大一辈。戈尔丹出生于1837年，1874年他被聘为埃尔朗根大学教授，是爱米·诺特父亲的同事。事实上，爱米是戈尔丹的博士生。据说她是他唯一的博士生，因为想成为他的学生可不容易。戈尔丹就学于柏林，他的数学研究非常规范且有高度的逻辑性，数学公式计算冗长，使用的是罗马字母而不是希腊字母。到了19世纪80年代，戈尔丹成为不变量理论界的世界首屈一指的专家。然而他无法证明一个关键定理，一个能使这一理论成为完整体系的定理。他只能证明在特殊的情况下这一定理成立，而不能证明其一般性。

希尔伯特登场。他于1862年出生在普鲁士的哥尼斯堡（即现在俄罗斯的加里宁格勒），1886年成为哥尼斯堡大学的私聘讲师。1888年他游访埃尔朗根，见到了戈尔丹。希尔伯特被那个不变量理论的著名问题深深吸引了。那个问题当时被称为戈尔丹问题，因为戈尔丹在这一理论上的权威性不容置疑。希尔伯特对此问题思考了好几个月，最后把它解决了！

希尔伯特在1888年的11月份发表了证明过程，并立即把一份证明寄送给了剑桥的凯莱。已经68岁高龄的凯莱马上回复了他："我认为你已经发现了这个伟大问题的解。"戈尔丹却没有那么热情。尽管当时希尔伯特还不是哥廷根派，但他的证明却具有"地道的哥廷根"风格：简洁、精妙、抽象、直观，使用希腊字母而不是罗马字母。这不符合戈尔丹的柏林人情结。他轻蔑地说：那不是数学，那是神学。1886年在哥廷根获得教授职位的克莱因正好在那里，他对这一证明印象深刻，因此当下决定要尽快把希尔伯特聘为他的助手。



我不打算陈述这一定理[4]的证明，只介绍希尔伯特不久之后所给的少了感性但更易于理解的结果：零点定理（英文为Zero Points Theorem，但人们通常用其德语名字Nullstellensatz）。零点定理引入了簇的概念，而

且展示了与几何的一种简单的关联。

从历史上来说，这一关联有点儿奇怪，因为希尔伯特是在代数数论中而不是在代数几何中发现的这一定理。这是一个关于交换环结构的定理，应该完全属于环论，不过现在代数几何学家已经牢牢掌握了它。打开任何一本代数几何课本，你都会在它的前两三章中找到零点定理。因此，我给出它的几何解释时就不会感到心虚，但是我希望读者记住，它的确是纯代数定理，是属于环论中的定理。

那么这个零点定理说的是什么呢？考虑由三个未知量x、y、z构成的所有多项式的环。首先，提醒你一下，它的确是一个环：加法、减法、乘法总是可行，除法有时可行。还要记住，设其中的一个多项式等于零定义了三维空间的某个区域，通常是一个弯曲表面。（这里我使用的是通常的笛卡儿坐标，而不是齐次坐标。）例如，当(x, y, z)是以原点为圆心、$\sqrt{8}$为半径的球面上的点时，多项式$x^2+y^2+z^2-8$等于零。请发挥想像力，把这个多项式与那个球面联系起来。

现在，考虑这个多项式环中的理想。理想就是一个子环，是这个环中的一个环，它还有一个性质：如果把这个子环中的任意一个元素与父环中的任意一个元素相乘，其结果仍在这个子环中。

例如，考虑所有这样的多项式：$Ax^2+Bxy+Cy^2$，其中A、B、C是x、y、z的任意多项式（包括零多项式）。这是x、y、z的所有多项式这个大环的一个理想。多项式$(x+y+z)(x^2+y^2)$是这个理想中的一个多项式，而多项式$x^3+y^3+z^3$则不是这个理想中的多项式。

现在，我要介绍簇这个关键概念。更确切地说是代数簇。从几何上讲，它就是二维曲线概念或者三维空间中表面或“空间曲线”概念的扩展。事实上，一个簇是某个多项式或一族多项式的零点的集合[5]。

所以，上面我提到的球表面是空间中多项式$x^2+y^2+z^2-8$等于零的点，它是一个簇。同样，方程为$x^2+y^2-4=0$的圆柱与这个表面的交集是一个簇（参见图13-3）。这个交集是由三维空间中两个半径为2的水平圆的圆周组

成的，其中一个圆周在xy坐标面上高为2的地方，另一个在它的下面距离为2的地方。这些圆是簇，是多项式对$x^2+y^2+z^2-8$和x^2+y^2-4的零点的集合。

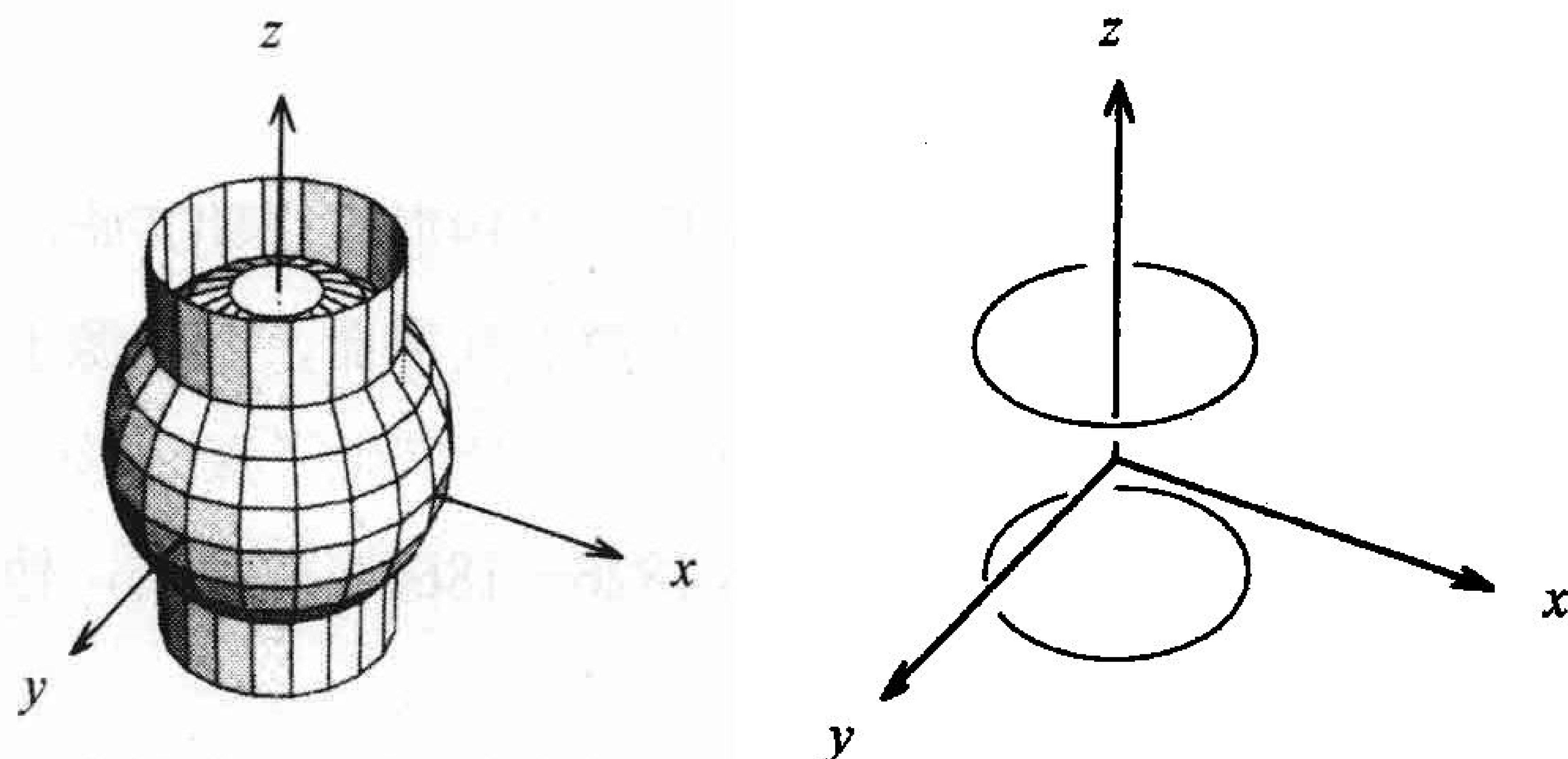


图13-3 两个多项式相交（左边）生成一个簇（右边）

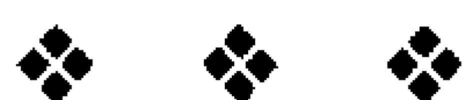
我刚才提到的那个理想是多项式的一个集合。因此，它定义一个簇：*所有使该理想中每一个多项式等于零的点*。这个簇到底是什么样的零点集合呢？答案是z轴。由这个理想定义的这个簇就是当$x=0$、$y=0$且z为任意值时得到的直线。

希尔伯特零点定理内容如下：如果一个多项式对于该簇上的每一点等于零（对于上面的情况，就是对z轴上的每一个点等于零），那么这个多项式的某个幂在该理想中。例如，多项式$7x-3y$对z轴上的每一个点都等于零，它不在这个理想中，但它的平方$49x^2-42xy+9y^2$在这个理想中。

当然，这里我已经相当简化了。没有必要一定是三个未知量x、y、z，可以是任意数目的未知量。我的例子中的簇也特别简单。做如此简化的最大过失可能是使用了这样的假设（我当时没有说，而是希望读者能够发现）：构成环的x、y、z多项式的系数是实数。事实上，它们必须是复数，零点定理对实系数多项式不总成立。[6]

在这方面，零点定理与代数基本定理相似。事实上，这两个定理关系密切，零点定理有时被称为代数几何基本定理。这种关联用所谓的零点定理的弱形式可以很好地表示：*对应于多项式环中理想的簇是非空的*

（除非这个理想是整个环）。构成一个理想的多项式一定有某些公共零点，因此这个定理称为零点定理。请与代数基本定理相比较，后者说的是一个未知量的多项式的零点集合是非空的。



1893年希尔伯特公布了他的零点定理。从19世纪中期开始，几何学还发生了三次革命。只有这最近一次，即第五次革命，是起源于代数学的，而第三次和第四次革命则在20世纪对代数学产生了深刻影响。

第三次和第四次革命都是由黎曼（1826—1866）引发的，他也许是最富有想像力的数学家。

在1851年在哥廷根发表的博士论文中，黎曼提出了黎曼曲面，也就是自相交弯曲面。在研究某些类型的函数时，它可以代替普通的复平面。

当把一个函数的变量放在复平面上时，黎曼曲面就诞生了。例如，复数−2i位于原点下方的负虚轴上。如果把它平方，就得到了−4，这个值位于原点右侧的负实轴上。可以这样想象：平方函数是把−2i沿逆时针方向旋转270°，使得它到达它的平方−4。

黎曼就是这样想象平方函数的。取完整的复平面，从原点出发朝向无穷做一条直线切口。抓住这个切口的上半部分，以原点为中心把它沿逆时针方向旋转。再把它向右拉伸360°。此时，它在伸展薄片的上面，而切口的另一边则在这个薄片的下面。让它穿过这个边（你必须得想象复平面不仅是可以无限伸展的，而且是由可以穿越自己的雾状物质组成的），并与原来的切口连接。此时想象中的图像有点儿像图13-4。这就是作用于ℂ上的平方函数的图像。

当从反函数的角度看它们的时候，黎曼曲面的威力就真正展现出来了。在数学上反函数比较棘手。取平方函数的反函数，即平方根函数。这个反函数的问题是任意非零数都有两个平方根。4的平方根是多少？答案是2或者−2。2和−2平方后的结果都是4。没有办法回避这个问题，但是，黎曼曲面却提供了一个更巧妙的方法来应付这一问题。

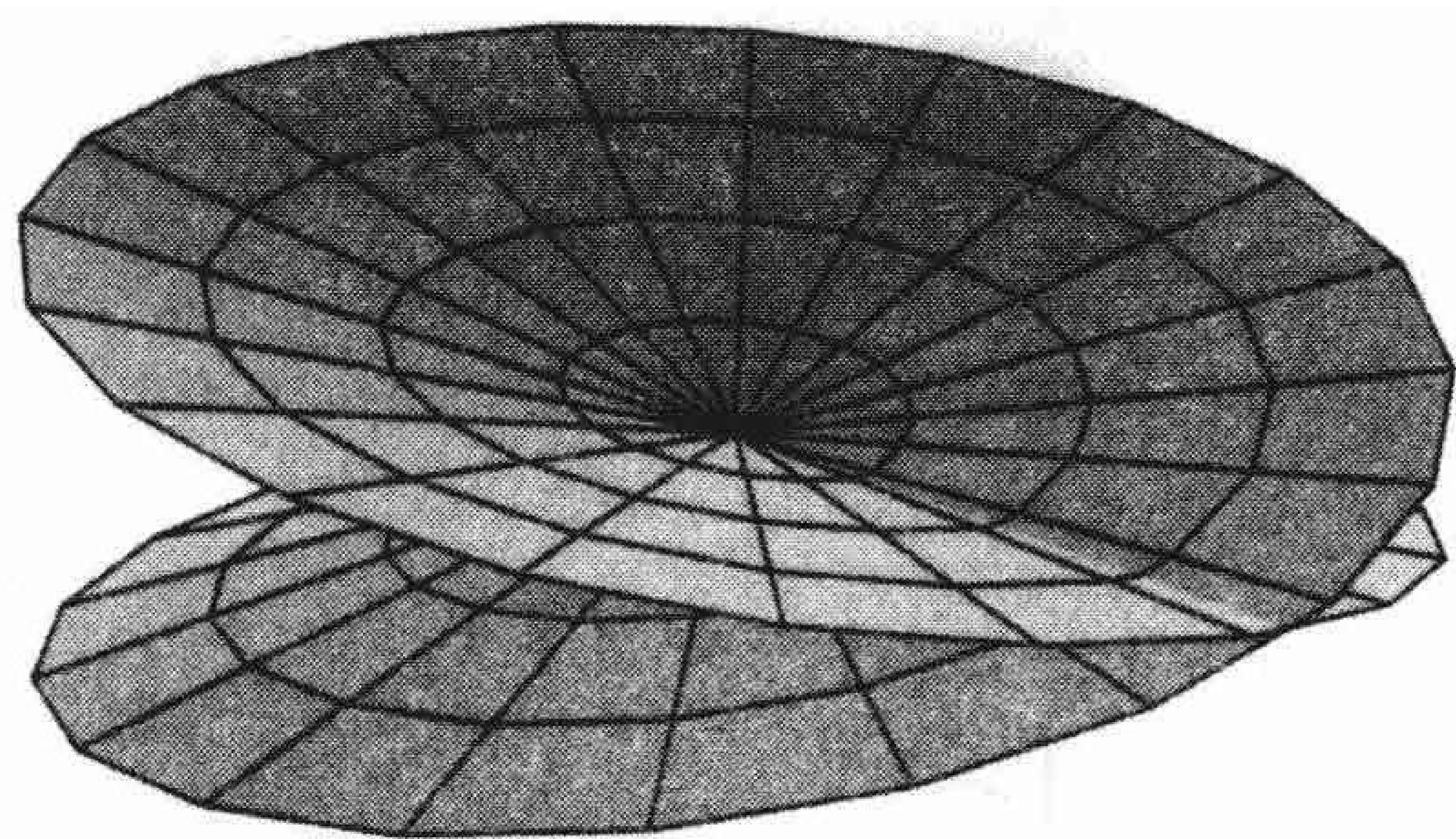


图13-4 对应于平方函数的黎曼表面

例如，–1的平方根是i或者–i。黎曼之前的数学家可能会用类似于图13-5的图像来表示这一情况。

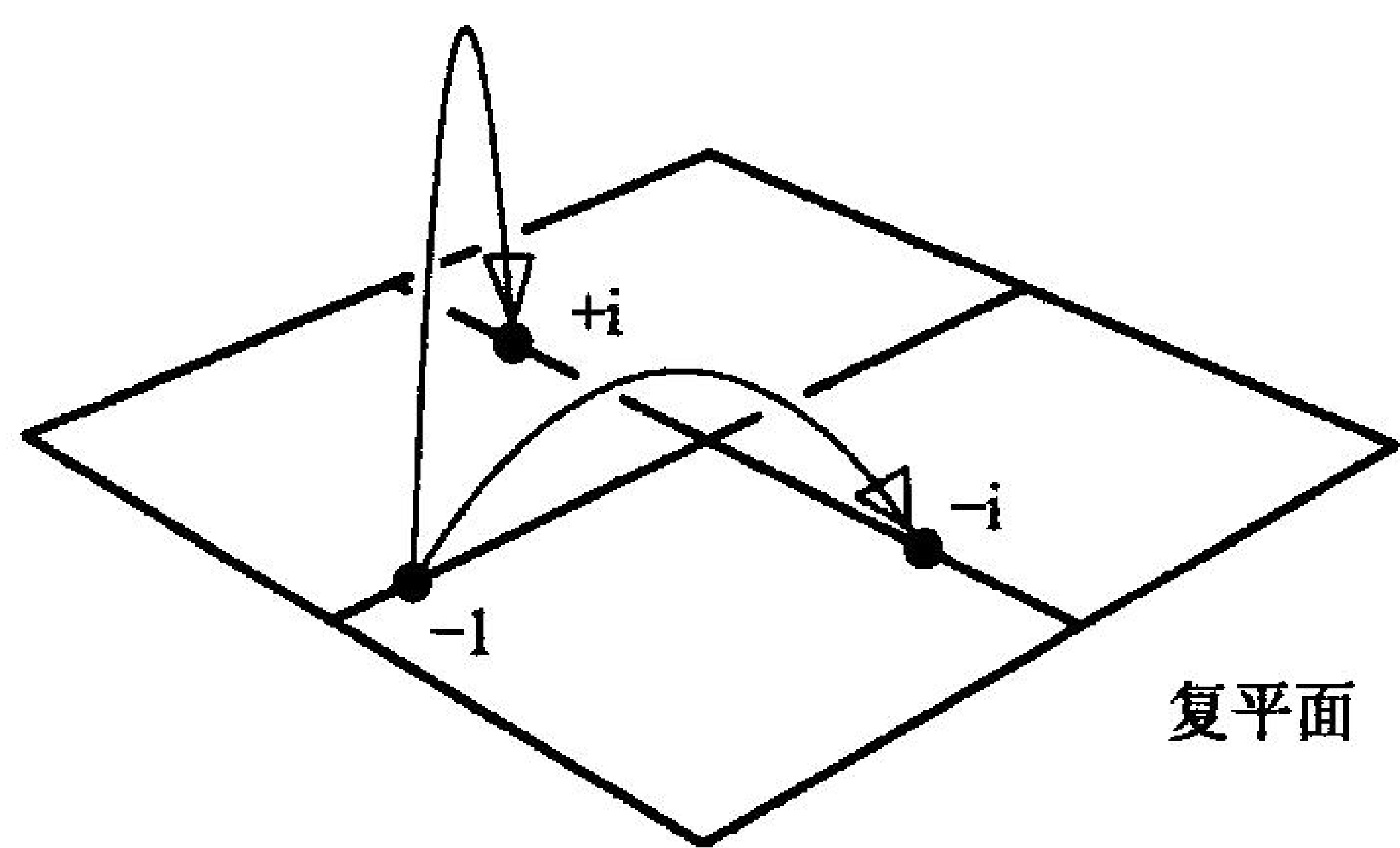


图13-5 数–1有两个平方根（黎曼曲线出现之前的观点）

然而，如图13-4所示的黎曼曲面把所有复数都成对地存储，一个在上，一个在下。（沿着折痕的复数除外，然而，折痕的位置是随机的，而且如果允许我使用实际需要的四维空间来画这幅图，我就可以使折痕消失。）

这表明图13-6也是一种表示平方根函数的方法。我通过考虑平方函数而给出的黎曼曲面可以看成表示平方函数的反函数（即平方根函数）的很好方法。–1有两个平方根，它们都在穿过黎曼曲面上两点的一条线上。

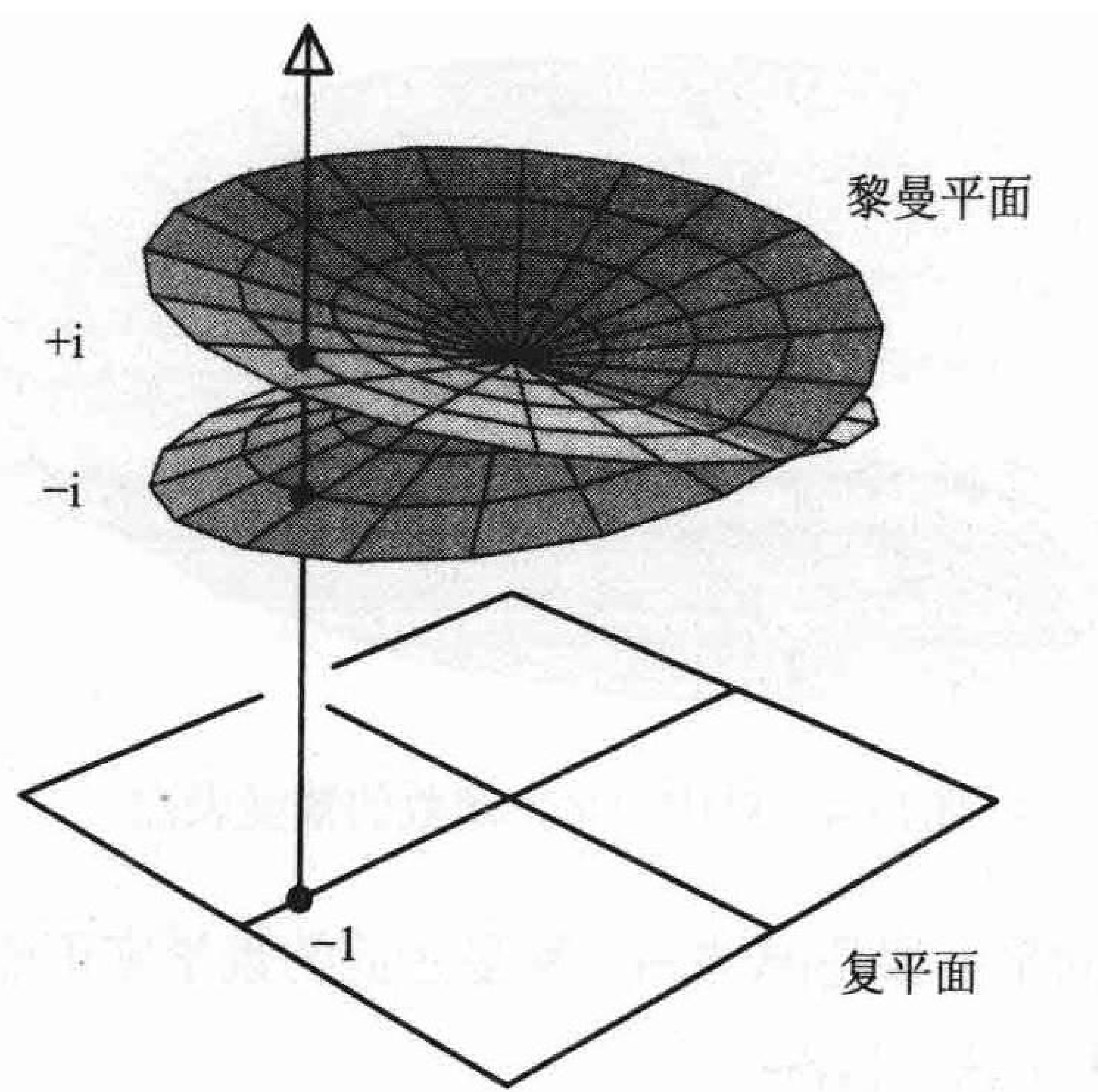


图13-6　数−1有两个平方根（黎曼曲线出现之后的观点）

第一次黎曼革命的重要性在于它为函数理论与拓扑搭建了一座桥梁。函数理论属于数学分析的范畴，而拓扑是几何的一个分支，黎曼曲面出现的时候，它刚呱呱落地。

下一章将更详细地讨论拓扑，这里只提及黎曼搭建的分析−拓扑这一桥梁，它打开了用20世纪发展起来的代数几何和代数拓扑的巧妙工具来研究函数理论的大门。其中一个核心定理是黎曼−罗赫定理，这个定理把函数的分析性质与对应的黎曼曲面的拓扑性质联系了起来。戴德金和韦伯于1882年联合发表了一篇论文，在这篇论文中，他们通过把理想理论应用到黎曼曲面找到了黎曼−罗赫定理的一个纯代数证明。这就是我在第12章中提到的那个结果。

（事实证明，在过去的140年间，黎曼−罗赫定理的更一般形式对数学家来说比其他任何定理都有用。）

黎曼并没有满足于引发了一场几何学革命，1854年他发表了自己的教授资格论文。这篇论文让数学界的人目瞪口呆，从而引发了第四次几何学革命。论文的题目是《几何基础之假设》，黎曼在其中构建了整个现

代微分几何，展示了60年后爱因斯坦用作广义相对论框架的数学。

使用黎曼曲面时，代数结果是间接的。黎曼的论文给出了20世纪关键概念流形的原型。所谓的流形就是一个“局部平坦”的任意维空间，也就是说在小范围可以近似看成普通的欧氏空间，这就像在日常生活中我们可以把地球的弯曲表面看成是平面一样（如图13-7所示）。这成为20世纪代数几何的关键概念。（流形的德文是mannigfaltigkeit，事实上它是由黎曼杜撰的，只是并不是在这篇论文中提出的。）

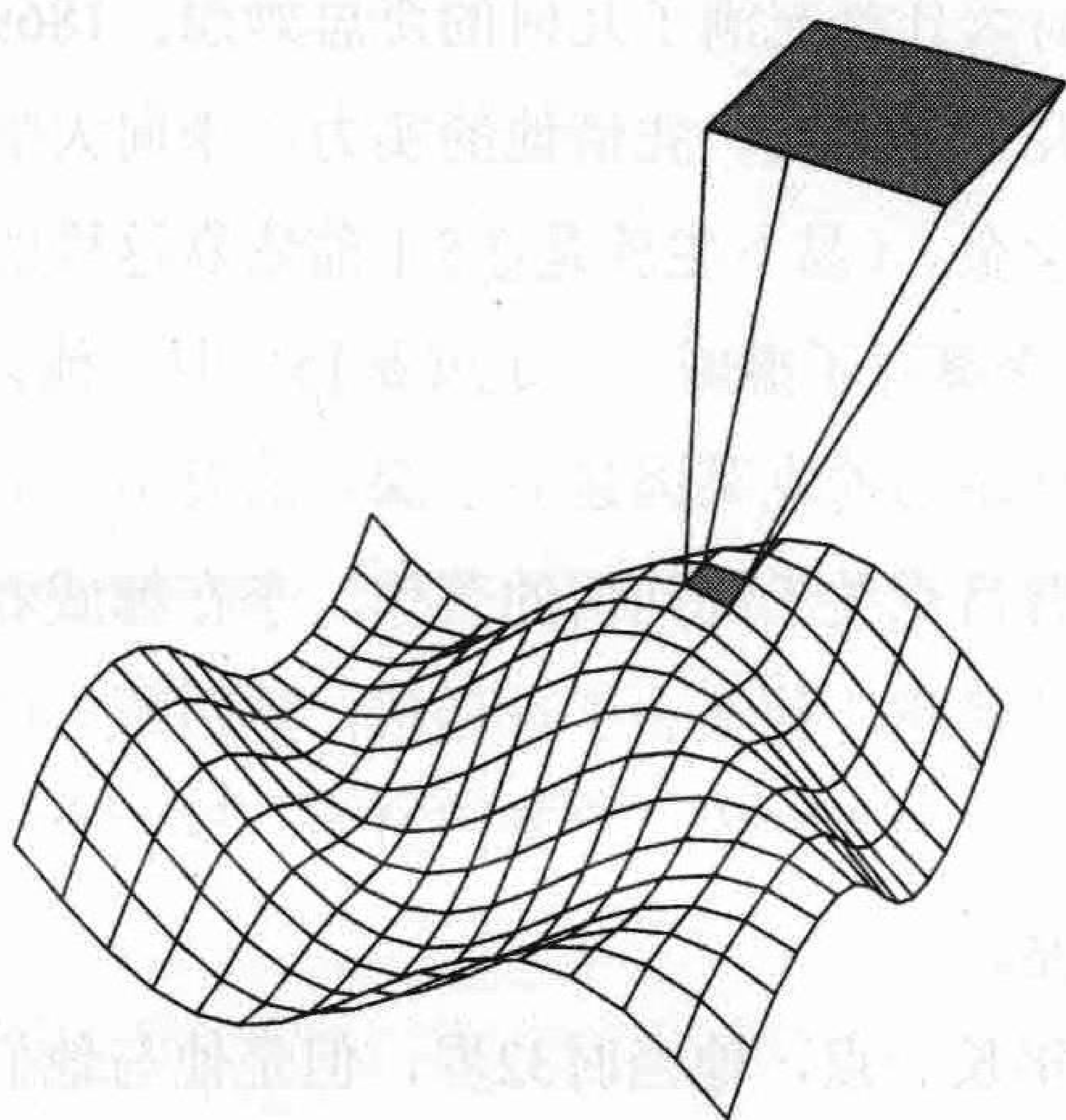
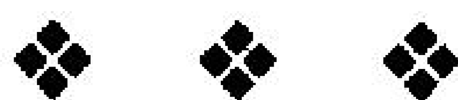


图13-7　一个流形可以有很多弯曲的地方，但是在每一点它是“局部平坦的”

19世纪第五次几何革命的成果虽然还延伸到了拓扑学、分析学和物理学领域，但它几乎是纯粹的代数革命。为了了解这场革命，我们不得不重谈一个话题，重访一个地方。这个话题就是群论，而这个地方就是岩石林立、海风呼啸的挪威海湾。



在第11章中，我提到了挪威数学家路德维希·西罗和1862年他在奥斯陆大学开设的关于群论的讲座。参加这一讲座的人当中有一位年轻的乡下小伙子，名叫索菲斯·李，当时只有19岁，长得像一个海盗：高高

的个子，满头金发，身材强壮，英俊而勇敢。他是一个远足迷，据说一天能走80公里。无论是在什么地形上跋涉，这个成绩都是非常不错的，更何况是在挪威这种冰川地形上，李堪称非凡了。有数学界的人说，如果李在徒步旅行时赶上下雨，他会脱下衣服，把它们塞到包里。但在斯图豪格为李所写的那本异常详尽并且满篇都是敬佩之情的传记中却没有提到这个说法。

尽管现在李被写成是一名代数学家，事实上他却总是认为自己是几何学家。他的所有著作都充满了几何的奇思妙想。1869年，他自己出资发表了一篇射影几何的论文，凭借他的实力，李向大学申请了到欧洲数学中心游访的奖学金。（易卜生就是在5年前依靠这样的奖学金跟挪威说再见的。）1869年李离开了挪威，一走就是15个月。他去了柏林，在那里他与当时也在柏林游访的克莱因建立了深厚的友谊。克莱因在普吕克去世后帮助出版了普吕克关于线几何的著作。李在挪威看过这本著作，并深受普吕克思想的影响。他又去了哥廷根，然后再去巴黎。在巴黎他再次遇到了克莱因，当时李27岁，克莱因21岁，这两个年轻人参加了卡米尔·若尔当的讲座。

若尔当稍微年长一点，他当时32岁，但是他与他们两个人属于同一代知识分子。他上学的时候原本打算成为工程师，没想到却成为了出色的数学全才。在1870年的春天，他刚刚出版群论的第一本著作《论代数替换与方程》。若尔当的著作没有达到凯莱1854年的论文的一般化水平。他把群写成了置换群和变换群。然而，这本著作涵盖的范围很广，被认为是现代群论的奠基性著作。至于李对西罗1862年的讲座记住多少我们不得而知，但是似乎可以肯定的是，正是有了与若尔当在巴黎相处的这几周，群的概念才深深地印入了他和克莱因的头脑。

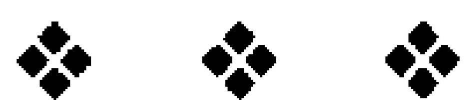
这一幸福而多产的数学伙伴关系到了7月19日被迫中止，当时法兰西帝国宣布同普鲁士王国开战。普鲁士人克莱因必须赶快离开法国。8月中旬，李也离开了巴黎，往南步行到瑞士，随身只携带了一个包袱。在距

离巴黎50公里的地方，他被当作德国间谍抓了起来，似乎是因为有人听到他自言自语，说的好像是德语。

宪兵检查李的包裹，发现了印有德国邮戳的几封信和笔记本，写满了奇怪的符号。李抗议说他是一名数学家。于是宪兵命令他解释笔记中的内容来证明自己的清白。斯图豪格的传记[7]中是这样记载这一事件的。

> 李应该怒吼："你们永远也不能够理解它！" [掌握了李理论的我们经常发出这样的声音……——笔者注]但是当意识到他的处境是多么危险时，据说他还是做了努力，他的开场白是这样的："先生，请考虑三个轴，x轴、y轴和z轴，它们互相垂直……"当他用手指在空中画出图时，宪兵们突然大笑起来，因此不需要进一步证明了。

在被允许继续前往日内瓦之前，李还是不得不在监狱里呆了一个月，靠读法译本的瓦尔特·司各特小说打发无聊时光。当他于12月回到奥斯陆的时候，发现自己成了19世纪挪威媒体关注的人物——一位曾被当作间谍抓起来的学者。1863年1月，他被奥斯陆大学聘为讲师和研究会员。不久后，克莱因也成为哥廷根大学的讲师。克莱因在普法战争中当了一段时间的卫生员，后来因生病而中断。而若尔当的书则被世界各地的数学家所传颂。19世纪70年代成为群论的第一个伟大的十年。而几何学的第五次革命也在这个令人惊异的世纪开始了：几何的"群化"。



在第11章中介绍了几种不同的群。它们都是有限群。每一个群都有有限多个元素。群可以是无限的，也可以是有限的。整数$\mathbb{Z}$以普通加法作为结合法则形成无限群。

几何学中包含丰富的无限群例子。在第11章中，我展示了二面体群D_4，一个含有8个元素的群，我们可以在原地旋转和翻转一个正方形来解释这个群。这是一个有限群，但是，如果解除上面的限制会怎么样呢？

如果我允许这个正方形平面以任意的方式移动到某个新位置，以任意角度旋转、翻转，又将会怎么样呢？（参见图13-8。）这一组动作是什么呢？

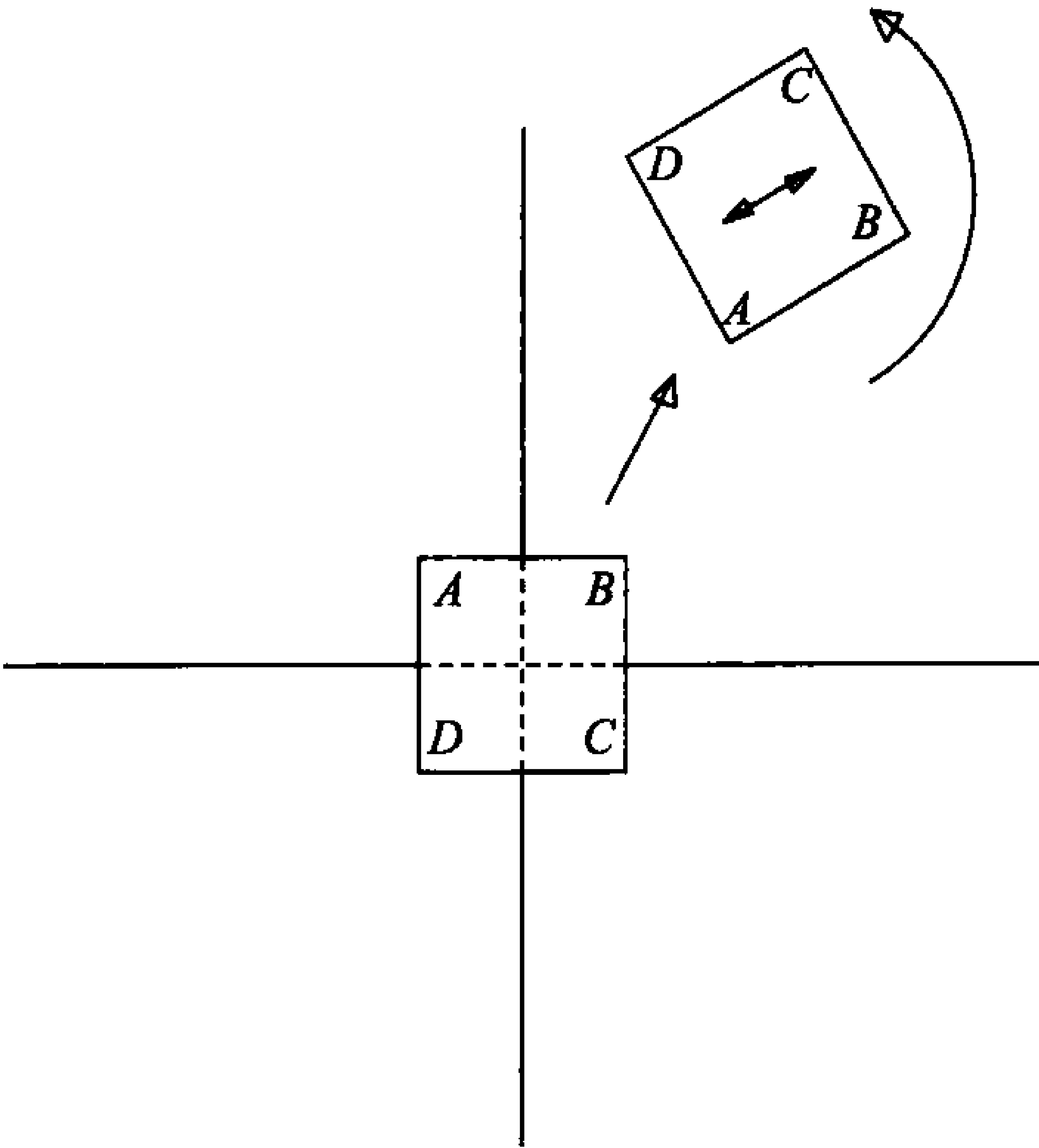


图13-8 一个等距

结论是这一组动作构成群！“移动这个正方形到一个新位置和新方向”的操作满足群运算的所有要求。如果你用一种方式移动它，然后再用另一种方式移动它，其合成结果就如同你用第三种方式移动它一样：$a \times b = c$。结合律$a \times (b \times c) = (a \times b) \times c$显然适用（“两个数相乘，然后再对相乘的结果做乘法……”）；什么也不做的“动作”充当恒等元，每一个动作都可以反过来做（“有逆元”），所以它是一个群。（问题：它是交换群吗？）

显然，这是一个含有无限多个元素的群。而且，如果想象一下把这个正方形画在一个无限大的透明片上，它在“原来”的平面上移动和旋转，你会看到这是一个整个平面的变换群。事实上，它的正确名字是欧几里得平面的等距群。“等距”一词在这里非常重要。它来自希腊语词根“等度量”，指的是“保持距离”的变换。如果两个点相距x，那么在任何等距变换下它们总是相距x，没有伸长，也没有缩短，任意两点之间的距

离是这个变换群下的不变量。

克莱因受到他与李及若尔当之间的交流的启发，有了一个绝妙的想法，19世纪前70年狂野扩散的几何学正是在这一想法的指导下统一到了一个巨大的组织原则之下。这一组织原则是：几何可以由保持其性质仍然为真的变换群来加以区分。二维欧几里得几何在刚才我描述[8]的平面等距群之下仍然为真。那个群的特征是某个不变量，在上面的情况下这个不变量是任意两点之间的距离。

我们能够从射影几何提取某个类似的群和特征不变量吗？能从罗巴切夫斯基和波尔约的双曲几何中提取某个类似的群和特征不变量吗？能从黎曼的更一般的几何中提取某个类似的群和特征不变量吗？是的，我们能够做到这一点。这里的群不太容易描述，但是我至少可以给出射影几何的不变量。显然，在射影几何中两点之间的距离不保持不变。三个点之间的两个距离的比率也不保持不变，这个结论不那么明显，但可以用图13-9加以说明。*AC*/*AB*的值是2，而其投影的比值是3。但是，如果如图13-10那样取四个点，并计算比率(*AC*/*AD*)/(*BC*/*DB*)，会发现在投影之下这个比率保持不变（在我的图中，它等于5/4），只是当把一个点投射到无穷远时，会出现稍稍复杂但也可以处理的情况。这个“交比”是投影不变量。

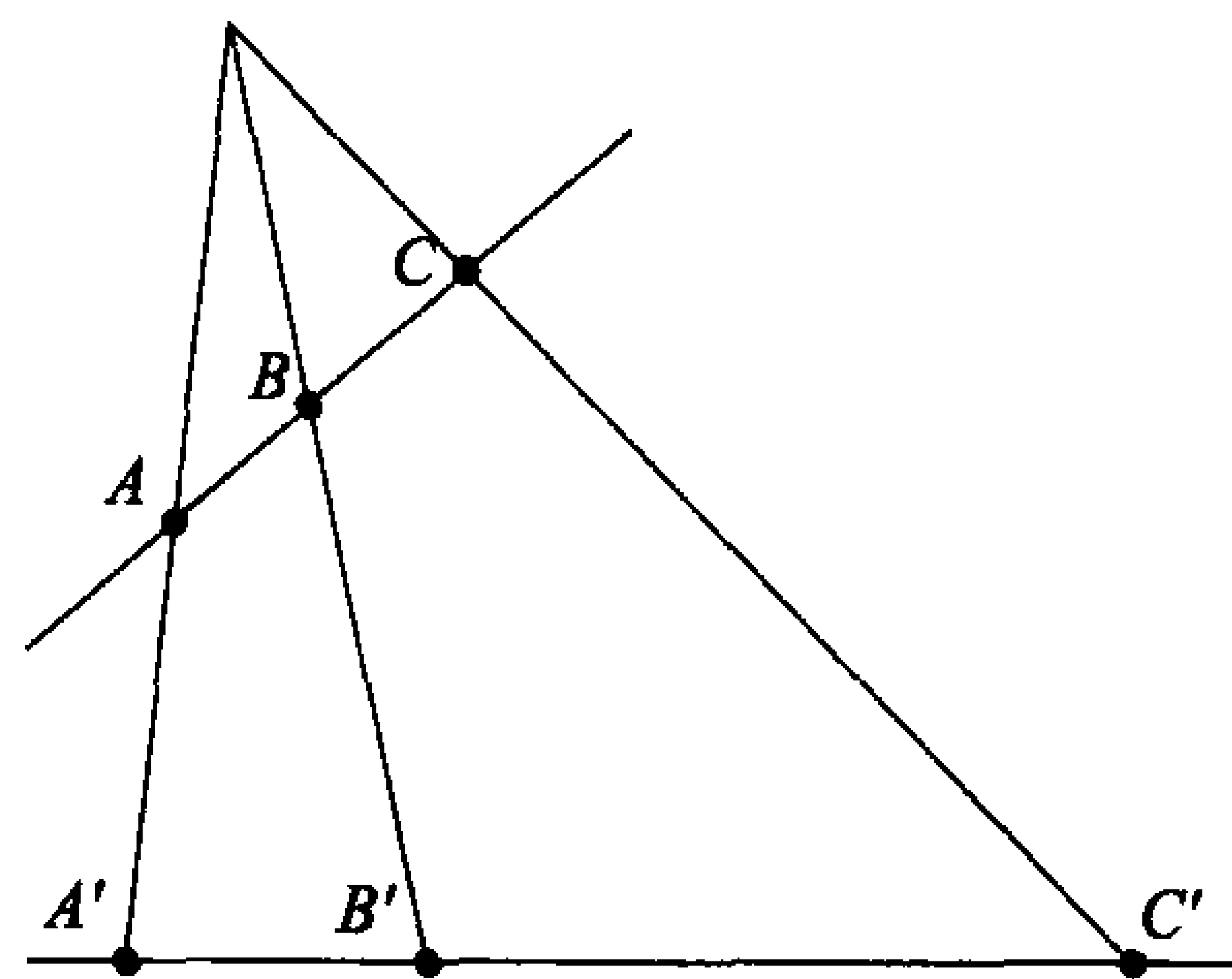


图13-9 *AC*/*AB*不是投影不变量

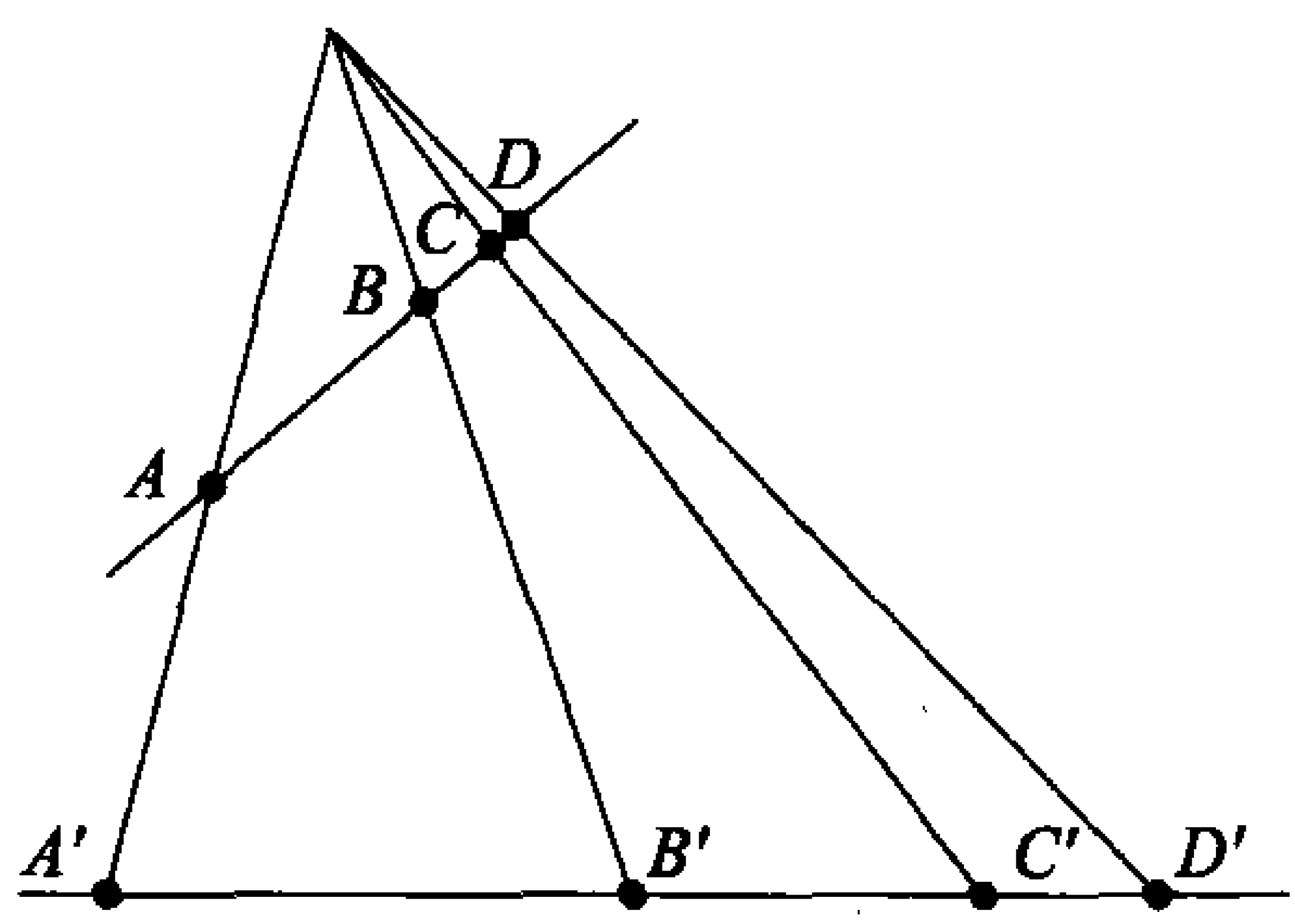
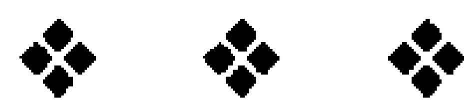


图13-10 （*AC*/*AD*）/（*BC*/*DB*）是投影不变量

1872年秋天，克莱因离开哥廷根去埃尔朗根担任教授。按惯例，新教授都要发表就职讲演，这是一种为其教授职称定调的演说，要展示出他要致力研究的领域。10月初，李与克莱因一起在埃尔朗根度过了两三周，帮助克莱因准备这份讲演稿。尽管最终克莱因没有使用这份讲演稿作为他的就职演说，但是他却把它作为一篇论文发表了，其题目是“探索新几何的几点考虑”。

这是不朽的伟大数学文献，被誉为埃尔朗根纲领。如果从论文要报告结果或解决问题的通常角度看，它不是一篇数学论文。它保有就职讲演的某些激昂的格调，只是克莱因在他的就职仪式上没有发表。在这份纲领中，克莱因给出了我上面提到的在群论和不变量理论之下统一几何学的思想。现在看来，这份纲领是战斗号角，号召数学家们积极行动起来去“群化”几何学。



在前文中提到的像平面上的等距群这类的几何变换群不仅仅是无限的，它们还是连续的，具有不可数无穷的性质。你可以将这个正方形移动1厘米，也可以移动1厘米的千分之一，或者1厘米的兆分之一。你可以把它旋转90°，也可以把它旋转1°的千分之九十或1°的兆分之九十。你可以不受任何限制地“切割”这些等距。它们甚至可以“无限小”。也就是说，在允许把这些种类的变换引入到类似于群的诸多事物之中时，我们就已经开启了微积分和分析进入群论的大门，反之也亦然。

克莱因本人并没有就此止步。在埃尔朗根纲领的末尾，他倡导连续变换群论，认为它与有限群论一样充满活力且将硕果累累。（他实际上说的是“置换”群。读者一定要记住，群论还远没有成熟。）

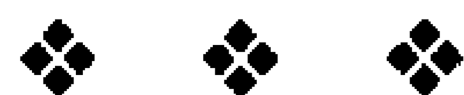
尽管李在促成这份埃尔朗根纲领形成过程中起着重要作用，但他还是认为这个想法太过于“雄心勃勃”了。1872年末时李已经是一位全职教授，这是挪威政府专门为他设立的数学教授职位。因为几乎不用承担教学任务，所以他全力投入引起他注意的一个问题。这是一个关于求解

微分方程的问题，在所谓的微分方程中，要求的未知量不是一个数而是一个函数，例如：

$$\frac{d^2y}{dx^2}+2a\frac{dy}{dx}+p^2x=0$$

我们可以用x的三角函数和幂函数的方式来求得y。李有处理这些方程的想法，这个想法与伽罗华处理普通的代数方程的方法有几分相似，但却使用这些新的连续群取代伽罗华理论中的有限置换群。

在与克莱因一起起草埃尔朗根纲领的时候，李对这方面的资料已经有了较深的了解，他认为变换群这个题目太宽泛而且混乱，无法按克莱因的建议进行严格的分类。一年之后他改变了想法，认为有可能构造出完全一般化的连续群论，它们的作用不局限于平面之内，而且可以作用于更一般的流形之上，可以得到高阶微积分的一些结果。李开始着手创建这个理论，今天这个理论[9]就是以他的名字命名的。



有了克莱因关于几何群化的埃尔朗根纲领，有了李的连续群论成果，有了希尔伯特1890年前后在环论领域的发现，19世纪代数和代数几何蓝图已几乎接近完成。在这里有一个国家我尚没有提及，但是在介绍代数几何时不提它是绝对不行的。

这个国家就是今天的意大利，继19世纪中期兴盛的民族意识的复兴运动之后，19世纪60年代意大利作为国家开始成型。因此，在统一国家之后，意大利人开始恢复他们光荣的数学传统，在19世纪末出现了一些优秀的学者：贝蒂、布里奥奇、克雷莫纳、贝尔特拉米。

黎曼对这些意大利数学家产生了深刻影响。在19世纪60年代早期，黎曼得了肺结核，工作受到影响，他便来到了气候适宜的意大利疗养。在这里逗留期间，他与几位数学家结为朋友。顺理成章地，这位张量分析（黎曼几何的现代形式）的传人不久就遇到了两位世纪末意大利数学家里奇和列维-奇维塔。

意大利人在几何方面特别强大。他们采用中世纪研究曲线和曲面的方法来研究他们自己感兴趣的东西，即某位数学历史学家所说的“现代经典几何”，并把它带到了20世纪。这就是出生于19世纪60年代和70年代的第二批意大利数学家的工作，这些数学家是卡拉多·塞格雷、圭多·卡斯泰尔诺沃、费代里戈·恩里克斯和弗朗切斯科·塞韦里。

然而，到了这些几何学家的工作成果成熟的时候，代数几何已失去了魅力。这不只是在数学时尚上的变化。到了20世纪初，逻辑和基础问题已经开始在“现代经典几何”中出现，这是庞斯列和普吕克早在100年前研究的几何。代数几何到了以希尔伯特和克莱因所预见的方式进行革新的时候。这次革新是我下一章的话题。在这里我只是让人们记住，在代数工具已经为代数几何在20世纪的转变做好准备之时，意大利人在保持代数几何活力的过程中所取得的成就。

我认为，“现代经典几何”结束的标志通常是指柯立芝1931年的教科书《论代数平面曲线》的出版。1873年柯立芝出生于马萨诸塞州的布鲁克赖恩[10]，他一生的大部分时间都在哈佛任教，从1927年一直到他退休的1940年，一直在哈佛科学院数学系担任系主任。他在著作中写了这样的题词：

> 献给所有意大利几何学家。

注　解

[1] 康德的思想是几何中分析与综合二分法的最根本源头。康德把分析事实和综合事实区别开来，分析事实的真理可以利用纯逻辑得到证明，不需要借助外界的任何引用，而综合事实可以通过其他方法而知道。在康德之前，哲学家都想当然地认为这“其他方法”的意思就是我们与这个世界接触的实际经验。然而，康德否认这一想法。在他的形而上学中，有一些事实不是分析的，也不是来自于经历的。他认为欧几里得几何的事实就属于这种，是分析事实但也不是来自于经验。这就是古希腊数学与19世纪初的“分析”几何之间的关联，不过我没有介绍这些联系的一些中间阶段。

[2] 噢，蔓叶曲线、螺旋线、外旋轮线、蜗牛形曲线和双纽线全都是特殊的曲线。例如，双纽线呈8字形状。尖点是曲线中的尖突，例如数字3在其中间有一个尖点。节点是曲线通过自己的地方，双纽线中间就有一个节点。凯莱曲线、海塞曲线和斯坦纳曲线都可以从给定的曲线通过各种操作得到。

[3] 实际上有很多“实现”二维几何齐次坐标的方法。一种方法就是重心坐标。在这个平面内挑出三条直线组成一个三角形。从任何一个点出发，画出到这个三角形三个角的直线。于是形成三个新三角形，每一个三角形都以你所选择的点为一个顶点，以原来的三角形的一条边为边。这些三角形的三个重心（适当做上标记）就起到了齐次坐标系的作用。重心坐标是莫比乌斯的重心坐标的过渡版本。在莫比乌斯重心坐标中，一个点是由三个重心定义的，这三个重心必须是在一个基础三角形的三个顶点上，目的是使选定的点是它们的质量中心。我们也可以在二维以上的空间中做类似的布局，当然代数很快就会变得更复杂。

[4] 这个1888年的结果应该称为希尔伯特基本定理，在任何一本好的高等代数或现代代数几何教科书中都是这么叫的。

[5] 迈克尔·阿廷在他教科书中说：“我不知道这个毫无吸引力的术语的出处。”我也不知道，然而我不觉得它毫无吸引力，当然不要与零点定理比较。杰弗·米勒介绍最早使用的数学术语时指出这是意大利几何学家欧亨尼奥·贝尔特拉米在1869年提出的术语。

[6] 我隐瞒了一件事，自19世纪中期以来，几何已经接纳了复数坐标。最初，在概念上很难适应这个事实，这就是我隐瞒它的原因。例如，如果你允许复数进入坐标或系数，那么直线就可能与其自身垂直！（在通常的笛卡儿坐标中，对于斜率分别是m_1和m_2的两条直线，当$m_1 \times m_2=-1$时，它们是互相垂直的。因此斜率是i的直线与自身垂直。）同理，在需要向刚刚结束了在分析课上与复平面搏斗的学生们介绍复数直线时，高等代数几何老师会欣赏这种说法吗？（“其坐标是复数的一维空间”，你会被弄糊涂的，一定会的。）

[7] 1861年古斯塔夫·罗赫（1839—1866）来到哥廷根在黎曼的指导下学习。他在黎曼去世四个月后就死了，非常年轻，只有27岁。

[8] 在这里我再次做了大量简化。事实上，正如克莱因知道的那样，如果你把膨胀的概念包含进来，一致地扩张或收缩整个平面，把一些图形旋转到另外的形状相同但大小不同的图形上去，此时欧几里得命题仍然为真。我忽略了这些复杂的内容。想要进一步了解这些内容的读者可以参考考克斯特于1961年出版的经典教科书《几何》第5章。

[9] 简言之，李群就是具有重要的“滑顺”性质的某个一般n维流形的连续

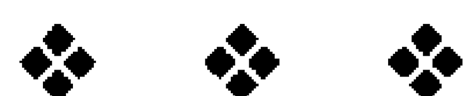
变换群。李代数是有一种做向量相乘的方法的向量空间。在李代数中向量乘法相当特殊，但是事实证明在某些高阶微积分的应用中它非常有用，而且很自然地产生自李群。

[10] 当被问到与这些上流阶层的布鲁克赖恩的柯立芝家族是否有关系时，出身较卑微的美国第30任总统柯立芝简短地回答说："他们说没有。"这一回答常为人所称道。事实上，所有美国的姓柯立芝的人都是沃特敦的约翰·柯立芝（1604—1691）的5个儿子的后代。这位总统是他的第二个儿子西蒙的第八代孙子，而这位数学家是第五个儿子乔纳森的第七代孙子，因此总统和数学家是相隔7代的表亲。朱利安·柯立芝的祖母是托马斯·杰弗逊的孙女。

第14章

无所不在的代数

大约从1870年开始——克莱因的埃尔朗根纲领可以看作一个转折点，整个数学界都开始重新理解代数。19世纪的代数学家发现的新数学对象（矩阵、代数、群、簇等）开始被数学家运用到工作之中。他们用这些新对象解决几何、拓扑、数论和函数论等其他数学领域的问题。我已经在第13章描述了一些几何代数化的内容。本章，我将介绍19世纪末及20世纪前期代数拓扑、代数数论和代数几何的进展。



先介绍代数拓扑。正如我在关于代数几何的数学知识中描述的那样，通常把拓扑介绍成为“橡胶几何”。想象一个二维曲面，例如球面，假设它是由某种可伸展的材料制成的。通过拉伸或压缩可以把这块橡胶转化成其他任意与球面“相同”的曲面，这就是拓扑考虑的东西。为了使拓扑具有数学的精确性，需要再制定一些规则，例如通过切割、粘合、“修剪”把有限区域变成无维点，或者就如这块橡胶是“薄雾”允许这个表面穿过自己，这些规则在不同的对象中稍有变化。在这里，这种不严谨但很形象的定义已经足够了。

直到19世纪末，拓扑都没有显示出与代数有多大关系。事实上，它早期的发展很缓慢。“拓扑”一词最初使用于19世纪40年代，是由哥廷根数学家约翰·利斯廷开始使用的。利斯廷的很多想法似乎来自于高斯，他与高斯关系很密切。然而，高斯从没有发表过任何与拓扑相关的东西。

1861年，利斯廷描绘了一种单面曲面，现在我们称其为莫比乌斯带（参见图14-1）。莫比乌斯四年后对这一东西做了描述，不知是什么原因，正是由于他的介绍，数学家们才开始注意这种单面曲面，所以它就被命名为莫比乌斯带。现在为其正名为时已晚，但是我还是如图14-1所示的那样标识该图，以彰显些许公正。（顺便提一下，如果你取我的图AG-3的有折痕的球，而且从其上切割下来一个小圆片，那么剩下的球面就与利斯廷带拓扑等价[1]。）

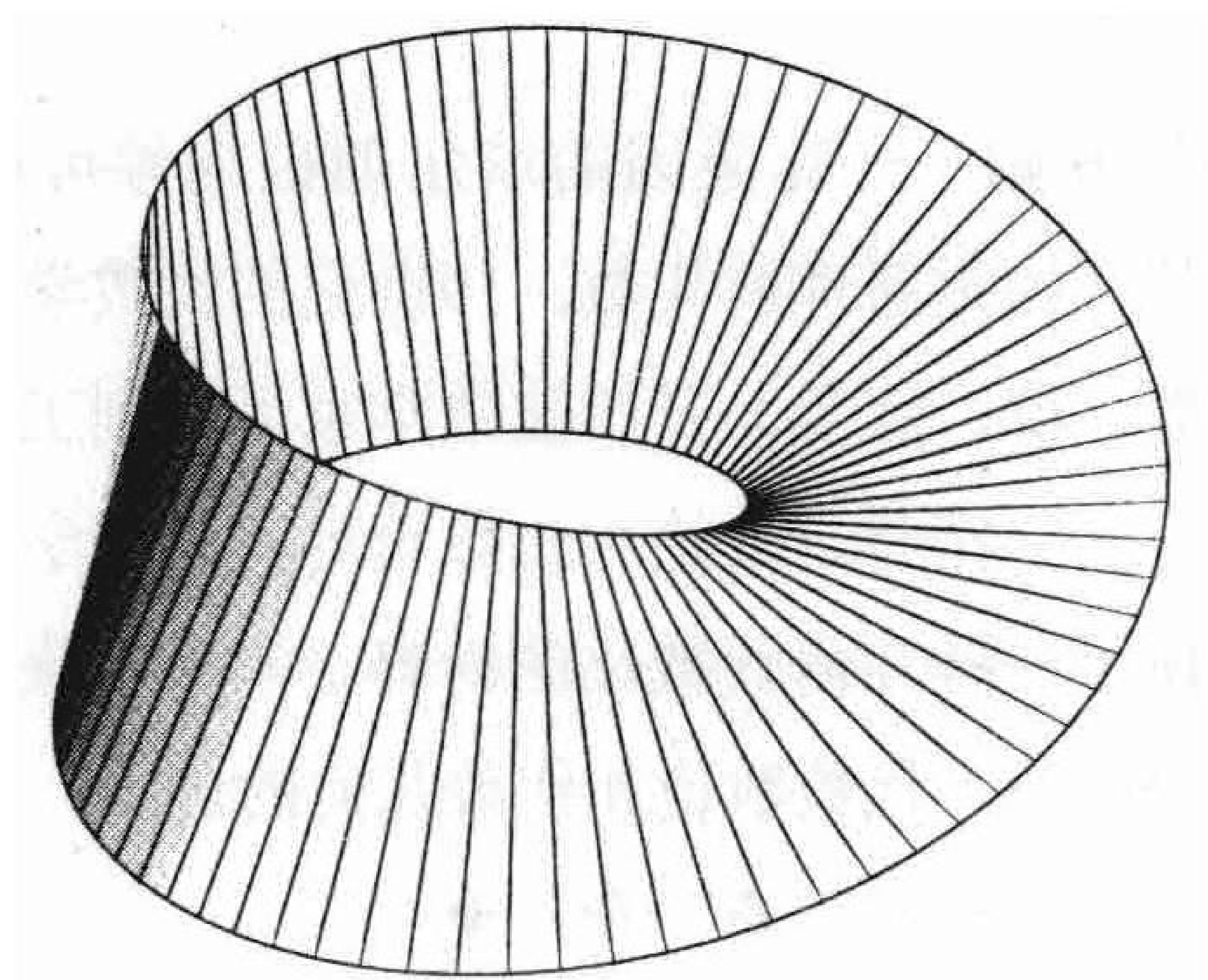


图14-1　利斯廷带

1851年，黎曼在他的博士论文中使用复杂的自相交曲面来帮助理解函数的做法是推动拓扑思想发展的另一个因素。在仔细研究这些黎曼曲面后，若尔当（参见第13章）提出的研究这些曲面的方法是观察嵌在其中的封闭环，看看会发生什么。在这里我用“曲面”来替代“空间”，使其更直观一些。

例如，想象一个球面。取球面上一点，从这一点出发，沿一个环行走，直到你回到这个点。对于你刚才走过的那条路径，如果对这个点不做任何非拓扑的行为，也从不离开这个曲面，它能一直收缩到这个出发点吗？能够平滑且连续地收缩吗？是的，它能。在这个球面的任何路径都如此。

在一个圆环面上就不能如此。图14-2的路径a或b都不能收缩到点P，只有路径c可以。所以，研究这些路径的确可以了解曲面拓扑的一些东西。

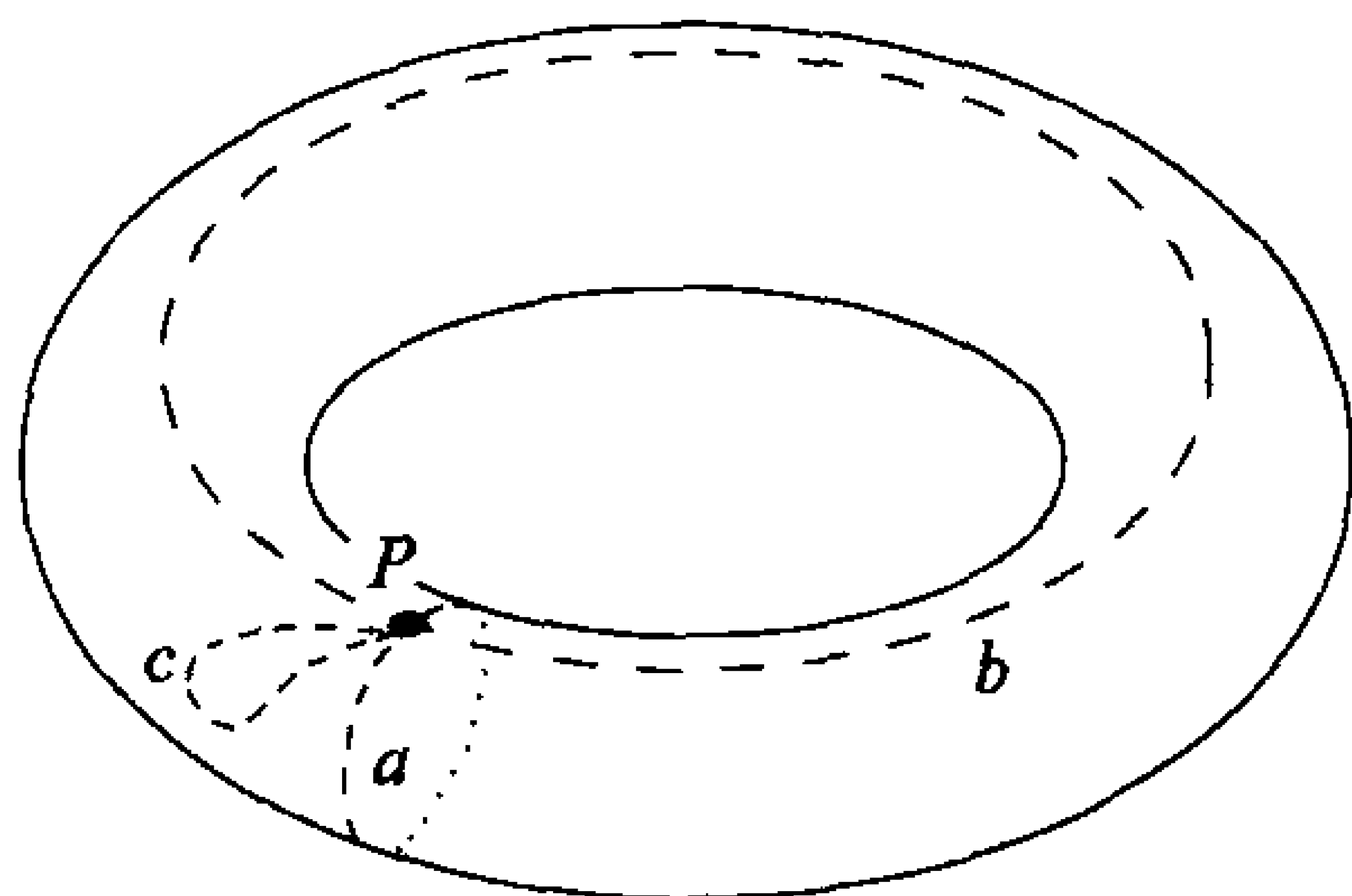


图14-2　圆环面上的环路径

1895年，一位优秀的法国数学家把这些想法代数化。这位数学家就是巴黎理工大学的庞加莱。庞加莱是这样说的。考虑一个曲面上的所有可能的若尔当环，即开始且结束于相同点的所有路径。保持基准点固定不动，把所有环组成若干族，如果一个环能够平滑地变形为另一个环，那么这两个环就属于同一个族，即它们是拓扑等价的。不论有多少个族，都要仔细研究。两个族的合成是这样定义的：经过第一个族的一条路径，然后再经过第二个族的一条路径。（选择哪条路径无关紧要。）

现在，你有了一个以环族为元素的集合，还有合成两个元素而得到另一个元素的方法。这些元素（即环族）能形成一个群吗？庞加莱证明它能够形成一个群，于是代数拓扑诞生了。

从这里开始还需要一小步就可以得到任意曲面的基本群的概念，你首先要摆脱你的环对任意特定基准点的依赖（它们不一定如我定义的那样是精确的若尔当环）。这个群的元素是这个曲面上的路径族。合成两个环族的法则如下：首先经过第一个族的一条路径，然后再经过第二个族的一条路径。

球面的基本群实际上是一个只有一个元素的平凡群。每一个环都可以平滑且连续地退变到这个基准点，所以只有一个路径族。

另一方面，圆环的基本群很奇怪——$C_\infty \times C_\infty$，它看起来有点吓人，但实际上不是这样的。取这个群的元素为所有可能的整数对(m, n)，且以加法作为合成法则：$(m, n)+(p, p)=(m+p, n+q)$。元素(m, n)相当于经过图14-5的a路径m次，然后再环绕b路径n次。（如果m是负数，你就沿相反方向走。按照下面的说法做也许有助于你理解这个群：如果你以某个角度从这个基准点出发，那么在结束环绕回到基准点之前，螺旋式地环绕这个圆环m次。）这些整数对在简单的合成法则下形成群$C_\infty \times C_\infty$[2]。

我说过，球表面的基本群是一个只有一个元素的平凡群。但是这一事实本身却不平凡。

事实证明以平凡单元素群为基本群的任意二维曲面一定与球面拓扑等价。现在，我们熟悉的嵌在普通三维空间里的二维球面在高维空间中有类似物。例如，有一个弯曲的三维空间"有点像"球，但是却是四维的，有时称它为超球面[3]。问题：基本群是平凡单元素群的任意三维弯曲空间与这个超球面拓扑等价的事实，在四维也为真吗？

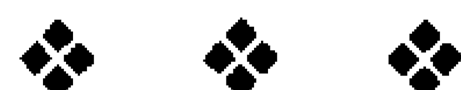
1904年，庞加莱提出了著名的庞加莱猜想，断言上面问题的答案是肯定的。2005年末，这个猜测还没有被证明，也没有被推翻。它也许是真的，在2002年和2003年发表的几篇论文中，俄国数学家格里高利·佩雷尔曼提出了一个证明，证明它为真。在我写这本书时，数学家们仍在评定佩雷尔曼的工作①。来自这些评价的非正式报告说，越来越多的人认为佩雷尔曼实际上证明了这个猜想。庞加莱猜想是七个千禧年大奖难题之一，解决其中任一个问题都将获得马萨诸塞州美国克雷数学研究所提供的100万美元奖金。如果佩雷尔曼的证明的确是正确的，那么他将拥有这100万美元。

当人们开始对一个数学理论品头论足时，它就开始活跃起来了。拓

---

① 现在，数学界已经认为庞加莱猜想被攻克了，其中我国数学家也做出了贡献。感兴趣的读者可以查证其中的故事。——译者注

扑伴随着庞加莱的著作《位置分析》的出版而活跃起来。在拓扑发展的最初几十年间，它被称为“位置分析”。直到20世纪30年代才普遍用“拓扑”来命名这一学科。我认为这要感谢莱夫谢茨，稍后我将详细介绍这个人。



在现代拓扑创始者庞加莱的猜想中存在一个奇怪的矛盾。

数学家认为拓扑实际上有两种含义：其一是几何含义，其二是分析含义。这里的分析指的是数学意义下的分析，即以函数、极限、微分和积分为研究对象的数学分支，这些都与连续相关。如果你回头看一下我多次提到平滑和连续变形，就会掌握这种拓扑意义下的联系。在某种意义上，从一个位置到另一个位置的移动，如果没有以平滑、连续、无限小增量地滑行等这样的基础概念为指导，或者说没有某些分析的思维方式，拓扑就没有意义。

用数学的行话说，分析的对立面是组合。在组合数学中，我们研究的东西都被标上1、2、3等，且整数之间不存在其他概念。因为相邻整数之间没有整数，因此也没有路径指引我们如何从一个整数过渡到另外一个整数。我们不得不跳过一个个空档。分析数学是连续平滑地通过连续的空间，组合数学是间断跳跃地从一个整数到另一个整数。

现在，拓扑应该是所有数学研究中最连贯的，它的橡胶平面可以平滑连续地弯曲和伸展。然而，最早出现的拓扑不变量却是表示孔数的整数，用来估量一个曲面内圆环形孔的数量，是由瑞士数学家西蒙·吕以利艾于1813年发现的。维数是另一个拓扑不变量（在拓扑意义下，你不能把鞋带变成一个薄饼，或把一个薄饼变成一块砖），它也是一个整数。甚至那些庞加莱揭示的基本群（如李群）也不是连续群，而是如“数学知识：数学和多项式”中定义的那样是可数的。尽管这些群是无限群，但是它们的成员可以标上1、2、3等等。对于连续群的成员不能这样做。所以，拓扑概念中的每一个对象都是间断的，而不是连贯的。

自相矛盾的是，庞加莱通过分析想到了拓扑，特别是通过他对微分方程的研究过程中的一些问题想到了拓扑。而他的结果以及他在《位置分析》中的所有思想都是组合的。对拓扑使用分析的方法（如今叫做点集拓扑学）对他没有吸引力。

这样的自相矛盾在庞加莱的重要的代数拓扑后继人身上更加明显，这个人就是荷兰数学家布劳威尔（1881—1966）。正是布劳威尔于1910年证明了维度是一个拓扑不变量。在现代数学中更重要的是他的不动点定理[4]。

布劳威尔不动点定理

n维球到其自身的任意连续映射都有一个不动点。

n维球就是实心单位圆盘（该平面上所有到原点的距离不超过一个单位的点）的概念及实心单位球（三维空间中所有距离原点不超过一个单位的点）的概念在n维空间中的扩展。在二维空间，这个定理意味着如果你把实心单位圆盘上的每一个点平滑地送到另外某个点，即把距离非常近的点送到距离同样非常近的点，那么总有一个点的位置没有改变。

不动点定理及其直接扩展理论有很多例子。例如，在你的杯子里平滑小心地搅拌咖啡。某一滴咖啡或者说某个分子就停止在它的出发位置上。（注意，从拓扑上说，杯中的咖啡是一个三维球。通过搅拌它，你就把咖啡的每一个分子从这个三维球的某个点X送到某个点Y。这就是我们说的“把一个空间映射到自身”的意思。）还有一个不太明显的例子：把一张纸放在桌子上，用笔在桌子上画出它的轮廓线。现在把这张纸弄皱，但不要撕破它，并把这已变皱的纸放入标出的轮廓里。这张变皱的纸的（至少）一个点一定在这张纸平放在桌面上你画轮廓线时的这个点的垂直上方。

布劳威尔拓扑中潜在的自相矛盾是他得到的结果与他的哲学本质相违背。对于一名普通的数学家来说这也许不是什么大不了的事，但是布

劳威尔却是一位非常特别的哲学数学家。他对这种形而上学的或者更精确地说是反形而上学的思想感到困惑，因为他的工作就是为数学寻找安全的哲学基础。

结果，他创立了直觉认知论学说。该学说要从人们在思考一系列相继的想法时的思维活动中寻找数学之根。布劳威尔说，数学命题不为真，是因为它对应于超越我们的实际感觉的某个柏拉图领域的更高级实体，我们的大脑还不能达到这个领域。它不为真，还因为它服从某个使用语言符号的游戏规则，就如布劳威尔时代的逻辑学家和形式主义者（如罗素和希尔伯特）所指出的那样。它为真是因为我们可以体验到它的真，这需要通过一步一步地进行某种适当的精神构建。根据布劳威尔的学说，构造数学的所有材料（非常粗略地说）不是来自超出我们感知的世界的某个仓库，它也不只是语言、纸上的符号和根据任意规则进行的操作。它是思考，是最终建立在我们对时间的直觉基础上的人类的生物学活动，它是我们人类本性的一部分。

这是直觉认知论的最粗略的要旨，它催生了大量的文学作品。了解这一哲学的读者会察觉到康德和尼采对它产生的影响。[5]

事实上，从任何意义上说，布劳威尔都绝不是这条思想线上仅有的创造者。类似于这样的思想贯穿现代数学历史，可以追溯到康德之前，至少是笛卡儿的时代。我想四维空间（第8章）的发明者哈密尔顿应该是直觉认知论者。1835年他在短评《论纯时间科学的代数》中试图把康德建立在“直觉”和“构建”之上且以几何为基础的数学思想引入到代数中来。

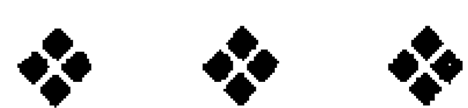
19世纪后期，克罗内克强烈反对乔治·康托尔把“完全无限”引入到集合论中，你可以称克罗内克为前直觉主义者。克罗内克说诸如$\mathbb{R}$这样的不可数集合不属于数学，数学没有它们也能发展，它们把多余且非必要的形而上学的包袱带入到数学之中，数学应该植根于计数、运算法则和计算。

这是一种思想流派，布劳威尔把它展现给20世纪的人，并把它传播给后来的数学家，例如美国人埃里特·毕晓普（1928—1983）。布劳威尔的学说称为“直觉主义”，毕晓普的学说则称为“构造主义”。人们今天了解到的思想就是“构造主义”，在美国它们的倡导者是柯朗研究所教授哈罗德·爱德华。爱德华教授2004年的著作《构造主义数学论说》中非常透彻地阐述了这一方法。（他在其他著作中也对此做了很好的解释。）

爱德华教授认为，因为拥有了功能强大的计算机，所以现在构造主义已经迎来了它自己的时代，而且，一旦人们的思维方式进行了适当的调整，那么1880年以来所取得的很多数学成果看起来都是误解。我没有资格评判这一预测，但也许是性情使然，我个人就觉得这位构造主义者的方法非常有吸引力，而且我是爱德华教授著作的铁杆粉丝，这可以从我在正文中多处引用他的话看出来。在第13章中我对手工制作数学模型和构造曲线的评论也可以看出我对构造主义的钟情。

总之，在他快30岁到30岁出头儿那几年中，布劳威尔在代数拓扑方面的研究一定是受到了他的哲学观的制约。十年后，他的同乡范德瓦尔登来到阿姆斯特丹向他拜师。范德瓦尔登在接受《美国数学学会通报》采访时说了下面的一段话。

> 即使布劳威尔最重要的研究贡献是在拓扑学领域，但是他也从来没有教过拓扑学的课程，而总是且仅是教授直觉主义基础。好像他自己不再相信自己在拓扑学方面的成果。因为从直觉主义的观点看，它们不正确。他根据自己的哲学将之前做过的每一件事情——他的最伟大的成果——都判定为不正确。他是一个非常奇怪的人，疯狂地热爱他自己的哲学。



下面介绍代数数论。这是一个不太容易准确说明的短语。因为存在被称为代数数的对象，所以在某些时候，代数数论表示对这些对象的研

究。代数数是可以成为某个一个未知量的整系数多项式方程的解的数。每个有理数都是代数数：119/242满足方程$242x-119=0$。由有理数、普通的算术符号、开方符号构成的任何表达式也都是代数的：$\sqrt[7]{18-\sqrt{11}}$ 满足方程$x^{14}-36x^7+313=0$。如第11章所述，根据伽罗华的研究，很多五次方程和更高阶的方程有解，但是不能用上面的方法表示，根据定义，它们的解是代数数。反之，很多数不是代数数，π 就是最著名的非代数数。（非代数数的数称为超越数。第一个证明 π 是超越数的人是费迪南德·冯·林德曼，他于1882年给出了证明。）代数数的理论体系很庞大，其基础是高斯和库默尔的工作，见第12章。你可以把上面所说的这一理论称为代数数论，数学家们经常这样称呼。

另外，像群这样的现代代数概念已被证明对解决传统的普通数论问题非常有用。其中最著名的问题，就是椭圆曲线上有理点的定位问题。这一问题听起来不太好解决，事实上，这条线路可以直接回溯到丢番图和他关于求诸如方程$x^3=y^2+x$的有理数解的问题（参见第2章）。如果画出这个方程相对于x的y的图像，就会得到数学家所说的椭圆曲线。丢番图的解对应于这条曲线上x、y坐标为有理数的点。有这样的点吗？有多少？我如何定位它们？这些问题听起来似乎不太令人兴奋，但是事实上，它们把我们引入了一个魅力非凡、令人着迷的数学领域，而且还带来一个伟大的未解问题：波奇和斯温纳顿-戴雅猜想[6]。

就在庞加莱宣布代数拓扑的前后，代数数论的范畴内的另一个话题是在环和域中发现了全新的数。这个数学分支是由亨泽尔开启的，他出生于哥尼斯堡（这也是希尔伯特的故乡），比希尔伯特大25天。亨泽尔于柏林在克罗内克的指导下研究数学，后来在德国西部的马尔堡成为教授，在1930年退休之前他一直在这个位置上任职。他对代数的贡献是1897年左右发现了p进数，这是代数思想运用到数论的一个很好的例子。

p进数中的p代表任意素数，为了便于说明我选择$p=5$。在我一开始描

述五进整数时，读者可以暂时忘记五进数。你也许知道，有一个与任意大于1的整数n相关的“时钟算术”，在这个算术中只使用$0,1,2,\cdots,n-1$。例如取$n=12$，就是普通的时钟表盘，但12被0取代。如果在这个时钟上把9加到7上（等价于问这样的问题：“7点再过9个小时是几点？”），你会得到4。用通常的记法就是：

$$9+7\equiv 4(\bmod 12)$$

好的，列出5的幂5、25、125、625、3 125、15 625，依此类推。为上面每一个数建立一个时钟：一个是从0到4的时钟，一个是从0到24的时钟，一个是从0到124的时钟，依此类推。从第一个时钟表盘上随机取一个数，比如3。在第二个时钟表盘上随机取一个匹配数（提到匹配，我的意思是：这个数除以5之后余数必须是3，）所以可以从3、8、13、18和23中随机选一个，比如我选8。现在，从第三个时钟表盘上随机选出一个匹配数，这个数除以25后余数必须是8，所以可以从8、33、58、83和108中选出一个，这里我选58。按照这样的形式永远进行下去，就得到一个数列，我把这个数列都放在括号里，并称其为x。它是这样的：

$$x=(3,8,58,183,2\,683,\cdots)$$

上面这个数列就是一个五进整数，注意是“整数”而不是“数”。我马上介绍“数”。给定两个五进整数，有一种方法可以把它们加起来，即对每一个位置运用时钟算术。也可以把两个五进整数相减和相乘，但不能相除。这与规范整数系统$\mathbb{Z}$很像，在这个整数系中可以做加、减、乘运算，但除运算不总可行。上面的“整数”系是一个环，是五进整数环，通常用$\mathbb{Z}_5$表示。

有多少这样的五进整数呢？在上面数列的第一个位置中，我可以从0、1、2、3和4这5个数中取出一个。同样，对于第二个位置，我可以从3、8、13、18和23中选出一个。对于第三个位置、第四个位置等其他位置都是如此。所以可能的数是无穷个5相乘的结果。粗略地将它表示为5^{infinity}，这是一个不太恰当的表示法。

这样的数可以组成其他集合吗？考虑0和1之间的所有实数，用五进制取代十进制表示它们。例如，把实数1/π写成五进制的形式就是：0.124 343 243 444 234 241 323 423 032 200 423 010 342 002 4…。显然，这样的“小数”（类似于五进制小数的东西）的每一个位置都有五种填入方法。这正好像我们的五进整数！所以五进整数的数量等于0和1之间所有实数的数量。

这是一件非常有趣的事情。到此我们已经创建了一个同整数类似的对象构成的环，而且它们与实数的数量相同！事实上，有一个定义两个五进整数之间的距离的实用方法，事实证明两个五进整数可以任意靠近，这跟普通的整数不同，两个普通整数的距离永远也不会小于1。所以这些五进整数有时像整数，有时又像实数。

对普通整数的环 $\mathbb{Z}$ 可以定义“分数域”，即有理数域 $\mathbb{Q}$ ，同样，对于 $\mathbb{Z}_5$ ，也可以定义分数域 $\mathbb{Q}_5$ ，在这个域中不仅可以做加法、减法和乘法，而且还能做除法。这就是五进数域。

如 $\mathbb{Z}_5$ 一样， $\mathbb{Q}_5$ 也是脚踏两只船。在某些方面它像有理数 $\mathbb{Q}$ ；在另一些方面，它又像实数 $\mathbb{R}$ 。例如，它是完全的，这像实数 $\mathbb{R}$ ，不像 $\mathbb{Q}$ （或者 $\mathbb{Z}_5$ ）。所谓“完全的”是指，如果一个五进数的无限序列趋近一个极限，那么这个极限也是一个五进数。不是每一个域都是“完全的”。例如，$\mathbb{Q}$ 就不是完全的。取 $\mathbb{Q}$ 中下面的数列：

$$\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \frac{239}{169}, \frac{577}{408}, \cdots$$

对于序列中每一个数，分母是前面的数的分子与分母的和，分子是前面数的分子与分母的2倍之和。所以有29=17+12，41=17+12+12。这个序列的所有数都在 $\mathbb{Q}$ 中，但是这个序列的极限是 $\sqrt{2}$ ，它不在 $\mathbb{Q}$ 中。[7]所以 $\mathbb{Q}$ 不是完全的。

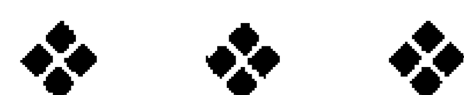
不过，我们可以把无理数加入到 $\mathbb{Q}$ 中使其变成完全域：$\mathbb{R}$ 是 $\mathbb{Q}$ 的“完全化”。或者说 $\mathbb{R}$ 是 $\mathbb{Q}$ 的一个完全化。p进数是另一种完全化 $\mathbb{Q}$ 的方法。

目前为止已经介绍了数论、代数和分析的一些概念：素数、环和域、无穷序列和极限。这就是p进数的美丽和迷人之处。亨泽尔的学生、马尔堡教授职位的后继人赫尔姆特・哈瑟把它们引入到20世纪的数学中来。哈瑟把p进数的基从普通素数扩展到一般数系中的“素数”，以此来扩展p进数。所谓一般数系中的“素数”包括我在域论知识中提到的那些数，以及高斯和库默尔在关于复数的因数分解及割圆整数的研究中所发现的那些数。

1934年，哈瑟离开马尔堡来到哥廷根。这里有一些内情，前一年，也就是1933年，纳粹掌握了德国的政权。哥廷根的所有犹太教授被迫离开，很多反纳粹的非犹太人（如第1章提到的奥托・诺伊格鲍尔）也被驱逐或因抗议而被迫离开。

在这一时期的种族分类体系下，哈瑟不是犹太人。然而，因为他有一位犹太人祖先，因此无论如何他都不是“纯”种人，所以不够资格成为党员。他似乎不是反犹太主义者，但是他是一位强烈的德国民族主义者，支持希特勒的种族政策。

在诺伊格鲍尔以及他的后任赫曼・外尔（这是另一名非犹太人，不过他的妻子有部分犹太血统）辞职后，哈瑟担任了哥廷根数学研究院的院长。他的动机似乎是，作为一名忠诚的民族主义者，他要保持德国数学的活力。而他必须接触的纳粹官员不喜欢他，一方面是因为他血统不纯正，另一方面是因为他的知识分子理想主义，这是当时的国家社会主义者不大欢迎的品质。然而，1945年希特勒失败后，他被英国驻军解除了大学的职务。在其快50岁时，面临再就业的局面，对此他没有抱怨。



最后介绍代数几何。我在第13章提到了“现代经典几何”形式的代数几何，被多塞格雷、卡斯特尔诺沃、安里奎斯和塞维里等20世纪初的意大利数学家所推崇。我已经说过，到了20世纪初，这种风格的几何开始遭遇动摇其基础的危机，一些很棘手的难题浮出水面，它们大都与曲

面和空间的“退化情况”相关，类似于我在介绍代数几何数学知识中描述的二次曲线的退化情况。到了1920年，这些难题越发棘手，已经阻碍了代数前进的步伐。

很明显代数几何需要革新，需要把它置于更坚实的基础之上，就如在19世纪从柯西到魏尔斯特拉斯等几代数学家对分析所做的工作一样。在20世纪30年代到40年代，代数几何的革新同样也是经过了几位数学家坚持不懈的努力才取得的。其革新的实质就是提升几何的抽象度。

我已经提到了克莱因的埃尔朗根纲领，其思想就是利用群作为组织规则把19世纪繁衍出来的射影几何、非欧几何、黎曼的流形（弯曲空间）几何等各种几何加以整理。

继克莱因之后，一旦数学家开始认为新几何是一个整体，是需要组织起来的思想集合，他们就开始注意所有几何共有的模式和原理。使几何更抽象、在任意特殊空间中都不引用点或线的想法已经扎根，19世纪后期的几位数学家，帕斯（在德国的吉森）、皮亚诺（在意大利的都灵）、魏纳尔（在德国的哈雷）努力尝试实现这种抽象。

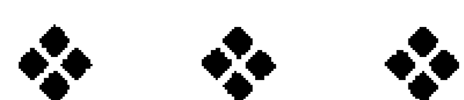
1892年，希尔伯特抓住了这一问题，当时他还是哥尼斯堡大学的无薪教师。他与几名同事去哈雷旅行，参加了魏纳尔的一个讲座。在这次讲座上，魏纳尔详细说明了他对几何学进行抽象的方法。返回哥尼斯堡时，希尔伯特一行人需要在柏林改乘火车。在柏林火车站等车时，他们谈论了魏纳尔的想法。希尔伯特做出了下面的评论：“无论何时都能用‘桌子、椅子和啤酒杯’来取代‘点、直线和平面’。”（把这一评论与第10章中引用的皮考克、格雷戈里和德·摩根在1830年到1850年对代数的评论对比一下。）

在接下来的六年里，希尔伯特没有在行动上穷追那句令人难忘的比喻，这段时间他终于获得了哥廷根大学教授的职位。然后，在1898年到1899年的冬天，他开设了很多讲座。在这些讲座中，欧几里得传统几何学被赋予了一个清晰、完备的抽象法则和公理，就如我在第11章中为群

给出的公理一样。希尔伯特说，公理所提到的对象可以是任何对象，但是他选择把它们说成是点、线和平面，其目的是为了使说明更加清晰。后来这些讲座内容被汇编成书，书名是《几何基础》。

这本书在数学家中广为流传，影响非常大。希尔伯特自己的数学研究随后转向其他方向，但是经常返回来审视几何。在1920年到1921年的冬天，他开设了“直观几何”的讲座，这一讲座涉及的范围比1898年到1899年的讲座更广，但不太抽象。这个系列讲座内容也被汇编成书，书名是《几何与想象》[8]，直到今天它仍很受欢迎。

希尔伯特对欧几里得几何公理化的处理给年轻的数学家带来了灵感。当然，前进的道路变得清晰尚需时日。有太多不同的观点，令人目不暇接：希尔伯特的公理化方法，克莱因1872年的群化纲领和1895年对拓扑的重写，希尔伯特对代数不变量的研究（零点定理和基本定理），以及这一领域中意大利几何学家利用19世纪中期的方法对所能接触到的曲线、空间和流形等进行的研究。



在对代数几何的最终革新过程中，有两个人脱颖而出：莱夫谢茨和扎里斯基。两人都是犹太人，都出生在19世纪后期开始日益强盛的俄国。

莱夫谢茨稍微大点儿，出生于1884年。尽管莫斯科是其出生地，但是他的父母是土耳其人，他们跟着老莱夫谢茨做生意而四处奔波。实际上莱夫谢茨是在法国长大的，他的母语是法语。作为布劳威尔的同时代人，与布劳威尔一样，他也在代数拓扑领域取得了成就。事实上，他甚至比布劳威尔更出名，有一个以他的名字命名的不动点定理。21岁那年他来到美国，1911年取得数学哲学博士，之前他在企业的实验室工作了五年。实验室工作的后果就是在一次电力事故中，他的双手被烧掉了。他的余生都戴着假肢，再戴一副皮革手套。在普林斯顿授课时（从1925年开始），都是由学生把粉笔放到他手里，然后他开始板书。莱夫谢茨精

力充沛，说话刻薄，为人固执，有几分西尔维雅·娜萨在《美丽心灵》[①]一书中所塑造人物的性格特征。他对自己与代数历史关系的总结非常生动："我的命运就是把代数拓扑这把鱼叉叉入代数几何这条大鲸鱼的身体里。"

扎里斯基比莱夫谢茨小15岁，出生于1899年。出生于这一时期的俄国很不幸。实际上，作为犹太人，无论出生在欧洲的哪里都是悲惨的。第一次世界大战的混乱，1917年的革命，德国的入侵以及随后的内战，这一切迫使扎里斯基背井离乡。1920年他去了罗马，在那里他在卡斯特尔诺沃的指导下学习。卡斯特尔诺沃是意大利学派"现代经典几何"的领军人物，此时他和他的同事已经明白他们的方法无法再取得进展。当时的卡斯特尔诺沃50岁出头，他觉得是把火种传递下去的时候了，他要求扎里斯基学习莱夫谢茨的拓扑方法。

当时正是20世纪20年代中期，墨索里尼及其法西斯成员正在加强对意大利人民生活的掌控。1925年，扎里斯基在罗马获得博士学位。很显然一两年之内意大利再不是他所希望的那样的避难所。当时莱夫谢茨正在普林斯顿，而扎里斯基在卡斯特尔诺沃的鼓励下与莱夫谢茨建立了友好的工作伙伴关系。1927年在莱夫谢茨的帮助下，扎里斯基在巴尔的摩的约翰斯·霍普金斯大学得到一个选修课任教职位。两年后他成为该校研究所的研究人员。

在20世纪20年代末到30年代初，扎里斯基致力于用莱夫谢茨的现代拓扑思想研究他从意大利学来的"现代经典几何"。这一工作的成果就是1935年出版的著作《代数曲面》。

然而，在编写和研究这本书的过程中，扎里斯基开始意识到代数几何的前进方向不能只走拓扑路线，还要利用希尔伯特在《几何基础》一书中所提到的公理化方法，而且要运用诺特的抽象代数。（数学研究已经

---

① 《美丽心灵》是20世纪伟大数学家纳什的人物传记。——编者注

到了关键性时刻的想法，是20世纪30年代末很多数学家的心声。这正是我在引言中引用的赫曼·韦尔发表的那段评论的背景。）1937年初，扎里斯基自己设定的工作目标是重新研究代数几何的基础。

此时扎里斯基已经成为美国数学家，他于1945年到1946年这一学年在巴西的圣保罗大学做访问学者，负责一个一周三个小时的讲座课程。只有一个人参加了这些讲座，此人是比扎里斯基年轻一点儿的法国数学家安德烈·韦伊[9]。

韦伊出生于1906年，是一个犹太人，还是一位和平主义者。在来圣保罗之前，为躲避欧洲的战争，他与扎里斯基同一时间来到美国教书。他是一位非常有建树的知名数学家，扎里斯基在此之前至少遇见他两次，一次是1937年在普林斯顿，另一次是1941年在哈佛。然而，他们共同在圣保罗的这一年对两人来说都是格外多产的一年。

与扎里斯基一样，韦伊也决定利用希尔伯特和诺特的抽象代数重新研究代数几何。而且他扩展了代数曲线、曲面和簇的理论，使得其结果在任何基本域都成立。基本域不仅有我们熟悉的实数域ℝ和（当时的）“现代经典几何”的复数域ℂ，还有诸如我在域论知识中提到的有限域等。这开启了代数与素数及一般数论之间的关系。韦伊的工作是现代数论代数化的基础。没有这一工作，安德鲁·怀尔斯就无法在1994年证明费马大定理。

19世纪出现的各种思想的溪流到此将要汇集到一起，形成对几何学的新认识，这一认识以抽象代数为基础，结合了拓扑学、分析学以及关于曲线和曲面的“现代经典”思想，甚至结合了数论。希尔伯特的零点定理和诺特的环、普吕克的线和李的群、黎曼的流形和亨泽尔的域都汇合在代数几何的统一概念之下。这是20世纪代数取得的伟大成就之一，同时也是一项引来颇多争议的成就。

注　解

[1] 在正确的拉伸和挤压下两个对象“相同”（即拓扑等价），可以用一个更

利落的、更时髦的行话来表述：同胚。我有点儿啰嗦了，为了表述上的简单起见，还是用“拓扑等价”这个词吧。

[2] C_∞被称为“无限循环群”。如果乘法对这个群适用，那么C_∞是由某个元素a的所有正次幂、负次幂和零次幂组成的：$\cdots, a^{-3}, a^{-2}, a^{-1}, 1, a, a^2, a^3, \cdots$。因为$a$的两个幂相乘等于它们的指数相加（$a^2 \times a^5 = a^7$），$C_\infty$就是适用普通加法 $\mathbb{Z}$ 中的普通整数群。因此，有时候你会看到把环面的基本群写成 $\mathbb{Z} \times \mathbb{Z}$，或者更谨慎些，写成 $\mathbb{Z}^+ \times \mathbb{Z}^+$，因为 $\mathbb{Z}$ 是一个环，而不是一群。

[3] 有时候也把超球面称为三球面。然而这个术语很难在头脑中产生很直观的感觉，至少对非数学家来说是这样的。“三球面”指的是弯曲的且定居于三维空间的普通球面的二维表面吗？还是弯曲的且定居于四维空间的超球面的不可见三维表面呢？对于数学家来说，三球面指的是后者，因为黎曼已经告诉我们，从一个流形自身内居高临下地考虑流形这一空间。然而，对于外行来说，通常把二维表面看成是由三维空间包围着的，前者就是这样的看似有理的想法。

[4] 拓扑学家海因茨·霍普夫（1894—1971）有一个经常与布劳威尔的FPT混淆的相关定理，这个定理告诉我们，此时在地球表面上的某个点处一定没有风，尽管是瞬间的。与此等价，想象一个球表面长满了一层短毛，你正尝试着向一个方向扫过这个球面。你会失败的。无论你怎么尝试，总是存在一个“漩涡点”，在这里没有毛倒塌。这也是一个定理，不过这个定理常被数学系的学生们不雅地称为“猫肛门定理”。（唉，我已经删除了一些有伤风化的句子了。）按此考虑的话，这个定理说的是：每一只猫必定有一个肛门。

[5] 然而在这里必须交待一下尼伯恩说的一件事：“康德的数学概念已经很陈旧了，说它和直觉主义者的观点之间有非常密切的关联很容易让人产生误解。不过一个有意义的事实是，如康德等人这样的直觉主义者是在直觉中寻找数学真理之源头，而不是在抽象概念的理性推断中寻找数学真理之源头。”摘自《数学逻辑和数学基础》的第249页。

[6] 这个猜想如庞加莱猜想一样，是克雷学院设置100万美元奖金寻找证明的问题之一。参见基斯·德夫林2002年的著作《千禧问题》，其中一共记述了七个问题。

[7] 证明这个序列的极限是 $\sqrt{2}$。根据形成项的规则，如果这个序列的某一个项是a/b，那么下面的项

$$\frac{a+2b}{a+b}$$

即

$$\frac{(a+b)+b}{a+b}$$

等价于

$$1+\frac{b}{a+b}$$

也即

$$1+\frac{1}{1+\frac{a}{b}}$$

也就是

$$1+\frac{1}{1+(\text{前项})}$$

如果这个序列收敛于某个极限数，那么项与项之间会越来越近，所以项和前项是几乎相等的。所以经过几亿项之后，它近似等于

$$x=1+\frac{1}{1+x}$$

如果运用一些初等代数方法，上面的表达式就是一个二次方程，它只有一个正根 $x=\sqrt{2}$。证毕。当然，这个证明不是非常严格，在理论上，它的缺陷是第二句话开始的假设“如果”。

[8] 这本书是希尔伯特与可因·沃什合著的，出版于1932年。两年后希尔伯特不再担任教学工作。

[9] 同李一样，韦伊也有被当成间谍抓起来蹲大狱的不幸经历，他的数学笔记和信件被怀疑是加了密的信息。事情发生在1939年12月的芬兰。被释放回法国后，他又因为逃避服军役而被抓了起来。

第15章

从泛算术到泛代数

说起近几十年代数领域的学术成果，通过下列美国数学协会颁发的柯尔代数奖获奖名单片段应该能见其一斑。（全部名单在因特网上可以查到。）

> 1960年：授予塞尔日·朗，嘉许他的论文“若干变量函数域上的非分歧类域论”；授予麦克斯韦·罗森利希特，嘉许他关于扩展雅可比簇的论文……。1965年：沃尔特·费特和约翰·葛瑞格·汤普森，嘉许他们合作的论文“奇秩群的可解性”……2000年：安德雷·苏斯林，嘉许他关于原动力上同调的工作……2003：授予中岛启，嘉许他在表示论和几何学方面所做的工作。

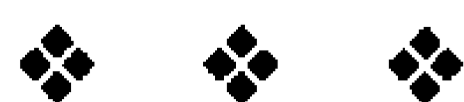
纵览上面的列表，可以看出我在本书跳过了一些代数内容，在此请求读者原谅。雅可比簇？非分歧类域论？原动力上同调？这些都是什么？

这就是现代代数，它以群、代数、簇、矩阵等一些重要的概念为基础。我希望我已经对这些概念给出了令人满意的陈述。即使是几个没有解释的术语也仅仅是脱离了19世纪基本思想一两步而已。例如，表示指的是群和代数的研究，其方法是利用我在第9章中给出的矩阵族来对群和代数进行建模。类域论是用来解决由于因式分解不唯一而引发的各种问题的非常一般化的现代方法，这些问题就是我在第12章中所陈述的令柯西和拉梅非常困扰的那些问题。可解性指的是群结构的事情，这一问题

可以追溯到方程的可解性问题，等等。

然而，代数已变得非常深奥却是事实，而且诸如原动力同调等话题对于非数学专业的读者来说不容易理解，我认为甚至是数学专业的人也不一定能够理解，除非他的专业就是这一领域。[1]代数也已经变得非常广泛，包含各种话题，在美国数学协会2000年使用的分类系统中有63个主题，它就占了13个[2]。

因此，现在我要实行作者的特权，只概述其中的三个话题以及最近几十年的某些人物，而不全面介绍最新代数。首先介绍范畴论，然后介绍亚历山大·格罗申迪克的生活与工作，最后介绍现代代数对物理学的应用。至于原动力同调，就等我写下一本书的时候再介绍吧。



20世纪后期，最受数学本科生欢迎的教科书是伯克霍夫和麦克莱恩的《近世代数概论》①。这本书第一次出版于1941年，它把20世纪中期代数的所有关键概念都非常清晰地整理到了一起，同时还为学生准备了数以百计的练习题来磨练他们的智慧。数、多项式、群、环、域、向量空间、矩阵以及行列式在这本书里都有介绍。我自己就是从伯克霍夫和麦克莱恩那里学习的代数学，我承认我的书受到了他们的影响。（不只如此，我还借用了他们书中的一些习题来帮助我解释。）

1967年，这本书的全新版本出版了，书名变成了《代数》。作者的顺序与原来相反：麦克莱恩和伯克霍夫。更有意义的是，陈述部分发生了变化。标题为《泛结构》的第四章是全新的内容，研究的是函子、范畴、态射和偏序集，这些术语在1941年的版本中根本没有出现过。另外还增加了长达39页的关于“仿射和射影空间”的附录。

一个本科生的课本需要在第一版出版后的26年做如此大的修改，这真有点令人惊讶。发生了什么？这些新的数学对象，假设从这些函子和

① 该书第5版的中文版和英文影印版已由人民邮电出版社出版。——编者注

偏序集我们大概能推测它们是什么，它们突然间从哪冒出来的？

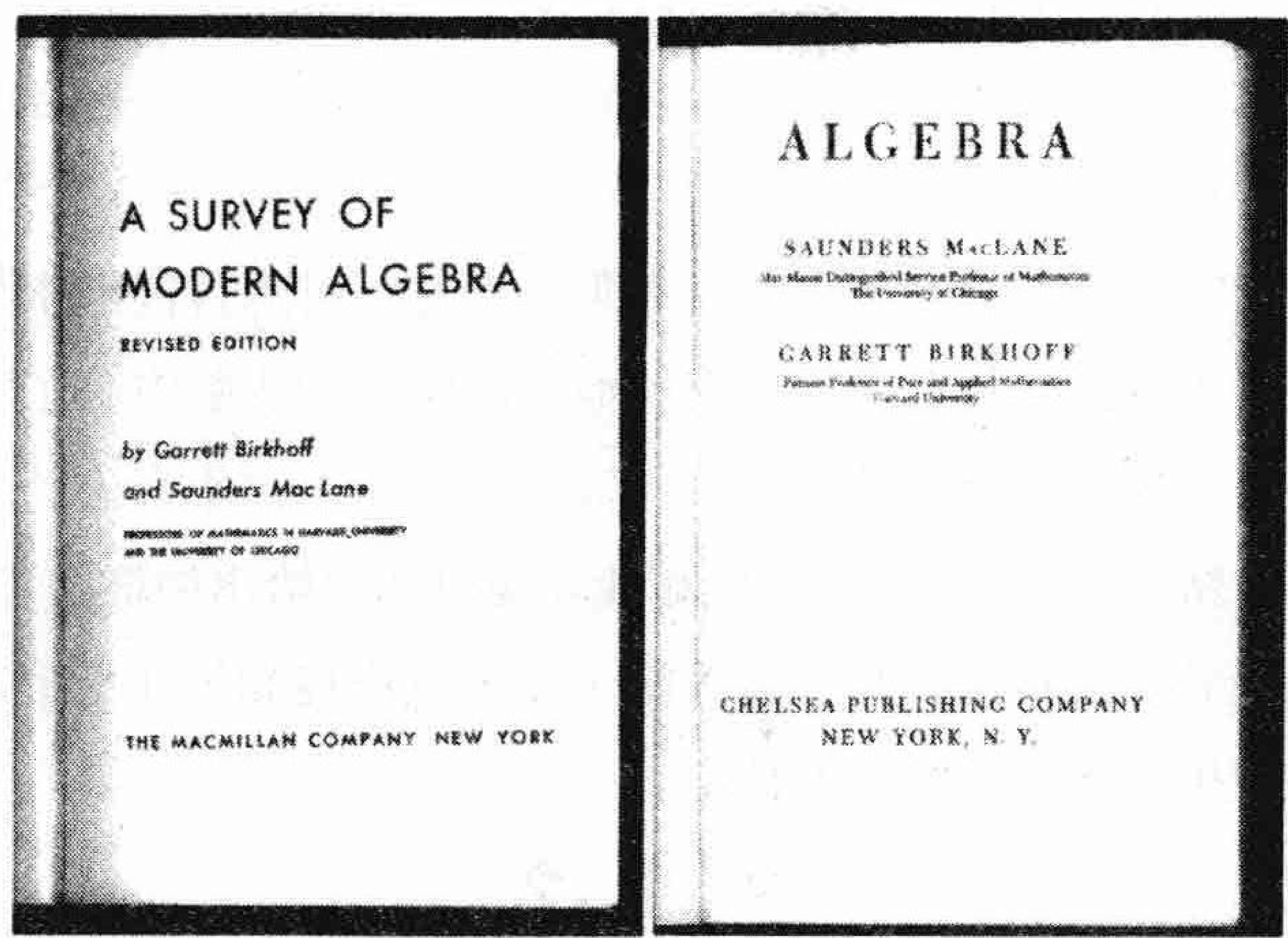


图15-1 伯克霍夫和麦克莱恩分别于1941年和1967年出版的两本书

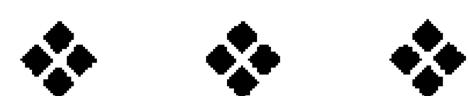
伯克霍夫（1911—1996）和麦克莱恩（1909—2005）都是20世纪30年代后期哈佛的教师。伯克霍夫的父亲从1912年到他去世的1944年也是这所大学的数学教授。爱因斯坦所描述的“这个世界的伟大的反犹太份子之一”正是这位老伯克霍夫，但实际上老伯克霍夫的偏见在当时当地也没有非常特别。[3]年轻的伯克霍夫于1936年被聘为哈佛的教员。麦克莱恩是康涅狄格州公理教会牧师的儿子，1934年到1936年在哈佛接受教育，1938年被聘为助理教授。

作为大学的代数教师，他们二人深受1930年在德国出版的一本书的影响。这本书就是范德瓦尔登的《近世代数学》，它利用完全抽象的公理化方法处理19世纪出现的新数学对象，这是在高层次数学上对这些新数学对象的第一次清晰的展示。范德瓦尔登给这本书增加了一个副标题“使用E.阿廷和E.诺特的讲座”。“诺特”就是我在12章中介绍的爱米·诺特。阿廷在纳粹上台之前一直是汉堡大学的优秀代数学家，纳粹上台之后他到了美国，曾在多所高校中教书。原计划是阿廷和范德瓦尔登联合

编写这本书，但是由于研究工作的压力阿廷退出了这一计划。范德瓦尔登的著作把新的数学对象群、环、域和向量空间整理到一起，并给出了它们的抽象公理化处理，就如希尔伯特、诺特和阿廷所研究的那样。

通过1930年的这本著作，范德瓦尔登把这种思维方法传播给数学界。通过1941年的著作，伯克霍夫和麦克莱恩把这一思想传授给本科学生。从此，术语“近世代数”在数学家和他们的学生的头脑中有了特别的意义。这一意义的本质就是如下研究代数的方法——用集合论语言表述的完全抽象且精确公理化的方法，就如我在第11章中给出的群的定义那样。

这是抽象的最新形式吗？是乔治·皮考克于1830年（见第10章）首次宣告的思想的终点吗？决不是！



1940年，正当《近世代数概论》准备出版时，麦克莱恩参加了一个在密歇根大学举办的代数拓扑会议。在这次会议上，他遇到了年轻的波兰拓扑学家塞缪尔·艾伦伯格，此人在一年前迁居到了美国，麦克莱恩对他出版的论文非常熟悉。他们成为了朋友，并于1942年共同发表了一篇关于代数拓扑的论文。这篇论文的题目是“群扩展和同调”，它研究的是同调，对此我需要简单介绍一下。

在第14章中，我用嵌入在流形中的环，即封闭路径的族，描述了流形的基本群。这个路径族的基本群是一个同伦群。通过扩展这些路径，即拓扑等价于圆的一维环，我们可以构建关联于给定流形的其他同伦群，方法就是把一维环扩展到二维、三维或更高维，分别拓扑等价于球、超球等“超环”。

同伦群非常有趣，也很重要，但在提供关于流形的信息方面，它们有一些缺陷。从数学观点看，它们有点笨拙。[4]

庞加莱发现了一个完全不同的与流形相关的群族。这个群族就是同调群。为一个流形构建同调群的直接方法就是用完全由单形体组成的近似流形替换这个流形。想象一个变形成为四面体（即以三角形为底的棱

锥，参见图15-2）表面的球面，你就会有概念了。现在你得到一个图，它是由零维的顶点、一维的边和二维的三角形面组成的。仔细研究经过这些点、边和面的各种可能方法，当从相反方向经过它们的时候，路径可以“抵消”（参见图15-3），这样就可以得到一个群族，通常分别将它们记为H_0、H_1和H_2。这些是同调群，它们合在一起称为这个曲面的同调。另外，还可以反过来进行整个过程，把顶点当作面，把面当作顶点，边仍当作边。[5]于是就得到了一个完全不同的群族，称其为上同调。

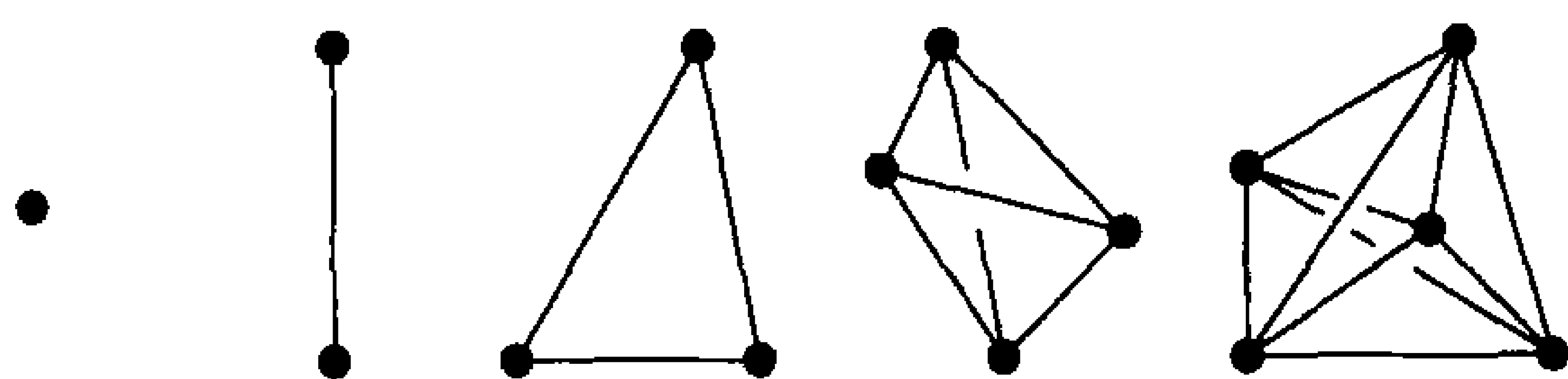


图15-2 从左到右分别为0单形体（点）、1单形体（线段）、2单形体（三角形）、3单形体（四面体）和4单形体[6]（五胞体）

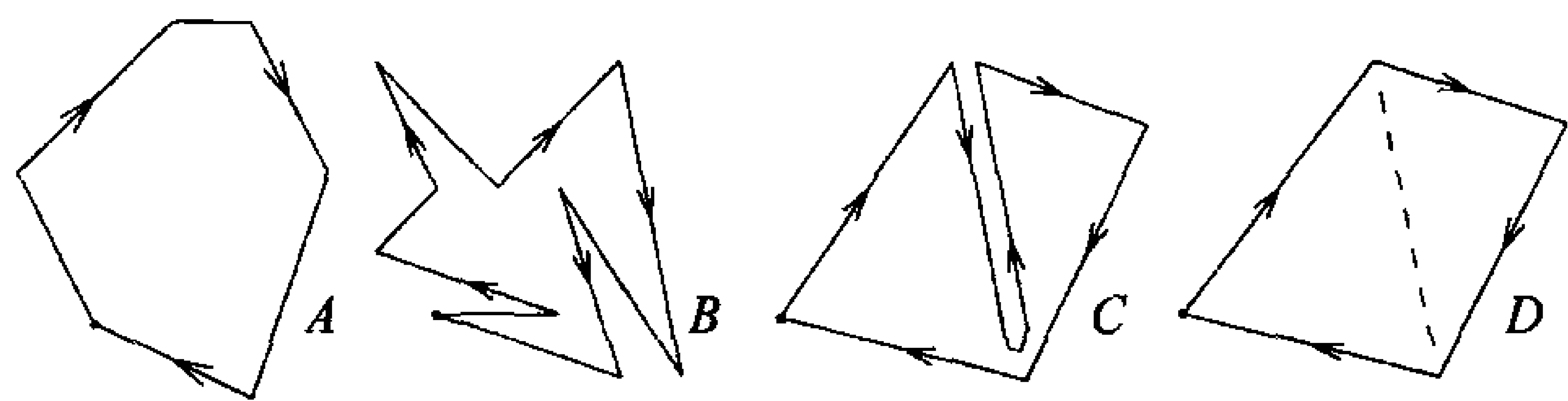


图15-3 路径A和B是同调等价的，路径C与D是同调等价的

我们可以对任意种类的流形在任意维度下做类似的同调。现在，三角形是可以放入任意区域的最简单的平面凸多边形。对于数学家，它是2单形体。3单形体是一个四面体，即三棱锥，有四个顶点和四个三角形侧面。4单形体是四维空间中的等价物，有五个顶点和五个四面体“侧面”（参见图15-2）。为了完整起见，我们称线段为1单形体，独立的点为0单形体。

所有流形都可以像这样“三角剖分”成单形体，当然，如果这个流形有洞从其中穿过，就像圆环面一样，那么就必须把几个单形体粘合到

一起形成一个“单纯复形”。一旦用这种方法对流形进行三角剖分，就可以找出这个三角形流形的同调，即单纯同调。这一同调和对应的上同调承载着这个流形的有用信息。另外，组成这个同调的群也比同伦群容易研究得多。（我顺便说一下，当制图师在测量地形时，就把类似的过程运用到实际的三角剖分中。参见下文。）

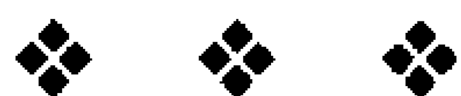
我们简单看了一下20世纪后期代数的一个关键概念，这是把代数对象附属于流形的概念。我提到的这个代数对象是群，就是我们在进行拓扑研究时出现的群。然而在同调论中，我们还可以把向量空间和模（参见第12章）附加到流形上。这为代数拓扑和代数几何开启了一个丰富的新领域。这个领域最著名的探索者是法国数学家勒雷、塞尔和格罗申迪克，稍后我再介绍他们。

这就是艾伦伯格和麦克莱恩1942年发表的那篇论文的背景。麦克莱恩（和伯克霍夫）刚刚完成了他们关于“近世代数”的著作，在当时已经普及的“近世代数”的启发下，艾伦伯格和麦克莱恩用非常抽象的方法研究了这个话题，从抽象度上看，他们的处理方法已经完全超出了我给出的少量描述，也超出了三角形和四面体。然而，在写这篇论文时，他们二人有可能想到了更高层次的抽象性。

在达到那种高层次抽象的三年后，他们又再次合作，论文的题目是“自然等价的一般理论”。就是在这篇论文中他们推出了范畴论。

我即将介绍的范畴论顺理成章地出自于同调理论。在以拓扑为原始背景发现了同调群以来的四十年间，人们发现它们与其他代数分支有着深层次的关联，特别是与我在第13章中描述的希尔伯特关于多项式环中不变量的工作密切相关。通过这种关联，黎曼已经建立了函数论与拓扑之间的关系。它们还与分析有关联，与研究函数及函数族的高阶微积分有关联。不久，在20世纪50年代，所有这一切发展成称为同调代数的研究领域。1955年艾伦伯格与他人合作编写了第一本关于同调代数的著作（是与伟大的代数拓扑学家嘉当合作的）。其中的一般化层次非常高，能

与范畴论自然地平行发展。



范畴论背后的一般思想路线是这样的。

诸如群、环、域、集合、向量空间和代数这样的代数对象是由元素（例如数、置换、旋转）和合成这些元素的一种或多种方法（例如加法、加法和乘法或置换的合成）组成的。当我们寻找把这些对象转换到或者说是“映射到”它们中的另一个对象或者它们自己的方法时，它们往往可以更清晰地展示自己的结构。（例如，回想一下，在关于向量空间的数学知识中我是如何把向量空间映射到它自己的标量空间的。同样，回想一下我对伽罗华理论的描述以及它与置换的关联。所谓的置换就是映射，即把一个解域映射到它本身且保持系数域不发生变化。）

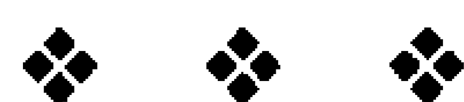
尽管对象是不同种类的对象，映射有不同的性质，但是，在所有情况下，贯穿元素、方法和转换的结构和方法大致相似。例如，考虑理想与其父环的关系（见第12章）以及正规子群与其父群（见第11章）的关系。这两种关系之间有某种纠缠不清的类似之处，那么，是否可以提出某些一般性原理，即代数结构的一般理论，使得所有这些对象以及我们将来可能遇到的其他对象都可以统一在某个超级公理的集合之下呢？统一在一种泛代数[7]之下呢？

艾伦伯格和麦克莱恩给出了答案：是的，这是可能的。为群或向量空间这样的数学对象族配备上对象间“行为良好的”映射族。这就是范畴，其中的映射称为态射。现在，你还可以再前进一步，建立一个范畴到另一个范畴之间的超态射。这种态射称为函子。

为了更好地理解这些概念，回头看一下我在第14章中对p进数的讨论。我构造了五进整数系统。然后，我相当从容地说：“就像对普通整数环$\mathbb{Z}$可以进一步定义‘分数域’（即有理数域$\mathbb{Q}$）一样，对于$\mathbb{Z}_5$，存在一种方法来定义分数域$\mathbb{Q}_5$，在这个域中你不仅可以做加法、减法和乘法，而且还能做除法。”上面过程中所隐含的技巧就是范畴论的函子的概念。

事实上，$\mathbb{Z}$ 不只是一个普通的环。它是一个相当特殊的环，这种环称为整环，其乘法可交换（对一个环来说这不是必要条件），它还有乘法恒等元“1”（对一个环来说这也不是必要条件），而且只能在a或b等于零或者二者都为零时$a \times b=0$（这也不是环的必要条件）。从 $\mathbb{Z}$ 得到 $\mathbb{Q}$ 的方法是从整环出发创建分数域。一般来说，从整环出发是一件可以做到的事情，因为我们能够构建一个从整环及其间的映射（或这些映射的子集）所构成的范畴到域及其间的映射（类似的子集）所构成的范畴的函子。

尽管我不打算深入讨论这些内容，但是我无法克制自己不提一下我喜欢的*忘却函子*。这个函子是这样的映射：它是从代数对象的范畴，比如说群的范畴，到（不考虑结构的）原始集合范畴的映射，“忘却”原来对象中存在的所有结构。



如此高的抽象层次上的数学有用吗？这个问题的答案要取决于问谁这样的问题。到2006年时，范畴论仍然有争议。在英语国家中，很多专业数学家在看到你在看范畴论时会皱眉摇头。只有少数研究生课程讲授它。“范畴”、“态射”和“函子”在迈克尔·阿廷长达600多页的权威性本科生课本《代数》之中提都没提。

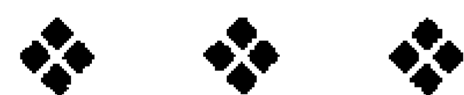
当20世纪60年代中期我自己也成为一名本科生时，最常听到的观点是，尽管范畴论可能是组织已存在知识的一种方便的方法，但是它的抽象层次过高，以至于难以建立任意新的理解。（我应该说，这是在英国，范畴论在此不被接受，认为是欧陆传来的，其实它来源于美国。）

无论如何，麦克莱恩非常喜欢他与艾伦伯格共同创造的理论。到了20世纪60年代中期，就在《近世代数概论》准备出修订版时，他重新改写了整本书，使之向范畴论方向倾斜。其他人也追随他，如果范畴论不能被普遍接受，当然也不能当作本科生的代数课程讲授，但它在数学世界中还是有很多铁杆支持者。它的追随者们信心十足地打趣着他们的最爱。威廉·劳威尔写了一本范畴在集合论的应用方面的著作，其开篇说：

“首先，我们剥夺了对象的几乎所有内容……”罗宾·甘地在《方塔纳现代思想辞典》中写到：“那些喜欢研究特殊、具体问题的人喜欢把[范畴论]说成是‘笼统抽象的废话’。”[8]

范畴论的支持者做了大量说明，其中一些超出了数学范畴而进入了哲学范畴。事实上，范畴论从一开始就在自我意识中带有一些哲学味道。单词“范畴”取自于亚里士多德和康德，而“函子”是从德国哲学家鲁道夫·卡尔纳普那里借来的，鲁道夫·卡尔纳普在他1934年的论文《语言的逻辑语法》中杜撰了这个词。

范畴论的哲学内涵本书不予介绍，只是在本章结束的时候我将对其内涵做一个简评。肯定有很多职业数学家利用这一理论取得了重要的成果。例如，亚历山大·格罗申迪克就是这样的一个人。



格罗申迪克是近代代数史中最具传奇色彩且最有争议的人物。大量关于他生活的文学作品不断问世，目前为止这些文学作品数量可能已经超过了关于他的数学工作的作品数量。其中，最可信而且最详尽的关于他的生活和工作的英文作品是艾琳·杰克逊的《仿佛来自虚空：亚历山大·格罗申迪克的故事》，这一故事在《美国数学学会通报》2004年10月和11月分两部分发表，这两部分合起来共计2.8万字。还有若干专门讨论格罗申迪克的网站。读者如果要了解这个人，可以看www.grothendieck-circle.org，里面也包含许多法语和德语内容，有艾琳·杰克逊写的上述两部分传记。

格罗申迪克的故事之所以引人入胜，是因为它是那种典型的讲述某个极具魅力“超现实”人物的故事：圣痴、疯子天才、隐居的虔心祈者。

首先说他天才的一面。格罗申迪克的辉煌年代是1958年到1970年。1958年，法国高等科学研究所（IHES）在巴黎宣布成立。它的所长是里昂·蒙项，一位有俄国和瑞士血统的法国商人，他认为法国需要一个与普林斯顿高等研究院类似的私立、独立的研究所。当时格罗申迪克30岁，

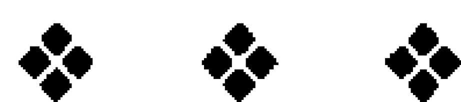
是IHES的创校教授。

IHES的私立和独立性因资金不足而受到影响。蒙项的个人经济来源远远不够支撑IHES的运营，从20世纪60年代中期开始，他就接受来自军方的资助。格罗申迪克是一名狂热的反军国主义者。当他不能劝说蒙项放弃军事基金时，就于1970年5月辞去了研究所的职位。

在IHES的12年间，格罗申迪克是一位数学新闻人物。他研究的领域是代数拓扑，而且他有能力把这一学科提升到一个更一般化的层次，使它成为数论、拓扑和分析的关键部分。

格罗申迪克继续前一代法国数学家让·勒雷的开创性工作。像130年前的庞斯列一样，勒雷也是在身为战犯期间得出了他一生中最重要的成果。勒雷于1906年出生，1998年去世，与安德烈·韦伊一样。身为法国军队的长官，勒雷在1940年国家沦陷时被俘，二战结束前一直被关在奥地利北部阿伦特斯泰格附近的集中营。勒雷那时的专业是流体力学。为了避免被德国政权强迫将自己的专业技术用于战争，他把研究兴趣转移到了当时他所了解的最抽象的领域——代数拓扑，把同调理论注入到了流形这个新领域。就是在这个领域，格罗申迪克和他的同代人法兰西学院的让-皮埃尔·塞尔继续研究前辈们的工作，成为领域先锋。

格罗申迪克是一位非常有魅力的老师，20世纪60年代那些学生对他表现出来的热爱之情有点像宗教信仰。他的数学风格不是所有人都能接受的，经常可以听到一些诋毁他的闲言碎语。然而，根据了解他及和他工作过的一些一流数学家（其中包括数学风格与他完全不同的数学家）的证词可以知道，在他辉煌的年代，他的数学创造力就如喷泉一样，散发着令人震惊的洞察力，总有深刻的推测，不断结出硕果。1966年，他获得了菲尔兹奖，这是数学优秀人材梦寐以求的最高级别奖项。获奖的评语是“以韦伊和扎里斯基的工作为基础，影响了代数几何基础的发展。”



我曾说过格罗申迪克是典型的圣痴和疯子天才。尽管我对他的很多

工作不能理解，但包括凯茨、阿廷、马佐尔、德林、阿蒂亚和沃沃斯基（其中有三个菲尔兹奖获得者）等人在内的诸多出色数学家对他的评价使我确信他是天才。神圣、愚蠢和疯子说的又是什么呢？

神圣和愚蠢指他如孩子般天真，了解格罗申迪克的每一个人都证实了这一点。这倒不是指他体格上像孩子。他（至少在年轻时）高大健壮、英俊潇洒，是一位出色的拳击手。在1972年发生在阿维尼翁的一次政治示威中，当时已经44岁的格罗申迪克打倒了试图抓他的两名警察。

在他二三十岁时，格罗申迪克全身心扑在工作上，远离尘世，对世间之事几乎一无所知。所有记载都说，似乎除了数学外他基本上不考虑其他任何事情。IHES教授路易斯·米歇尔回忆说，大约在1970年左右他告诉格罗申迪克有一个会议是由NATO（北约）主持的。格罗申迪克看起来很困惑。米歇尔问他知不知道NATO是什么？回答是“不知道”。

他忽视的范围也扩展到了他不感兴趣的某些数学领域，据说除了绝对抽象的代数研究之外所有其他领域他都不感兴趣。例如，他不觉得数值有趣。有时候数学家把57叫做“格罗申迪克素数”。57是3与19之积，因此这个数不是素数，那么为什么这么叫呢？故事是这样的，格罗申迪克参加了一个数学讨论，当时有一个人建议他们提出一个适用于所有素数的素数检测过程。“你是说一个实际的素数吗？”格罗申迪克问。这个人回答说：“是的，是一个实际的素数。”格罗申迪克说：“好吧，就取57吧。”

格罗申迪克自己可怜的童年毫无疑问都贡献给了这个他无法理解的尘世和数学。他的父母都是有强烈反叛思想的人。他的父亲是乌克兰犹太人夏皮罗，出生于1889年，他的一生几乎都处于无政府主义政治的阴暗世界之中，维克多·赛尔吉在他的回忆录中对这个世界做了描述。夏皮罗几次被关进独裁政治的监狱，为了逃避沙皇的独裁政治，他甚至试图自杀但未遂，失去了一只手臂。1921年他前往德国，做街道摄影师谋生，没有受雇于人，不违反他自己的无政府主义者原则。

格罗申迪克的母亲是约翰娜·格罗申迪克，她的孩子就随了她的姓。

约翰娜不是犹太人，出生于汉堡。她不顾自己的中产阶级教养和体统，独自到柏林生活，靠为左翼报纸自由撰稿谋生。格罗申迪克就出生在柏林，他的母语是德语。第二次世界大战爆发时，这一家人在巴黎。然而，在西班牙内战中，他的父母站在了共和党一边，所以被法国战时当局认为是潜在的危险分子。格罗申迪克的父亲被捕，并被用船押运到奥斯威辛杀害了。孤儿寡母在俘虏收容所呆了两年。之后格罗申迪克被带到了法国东部的一个小镇，在那里抵抗组织很强大。在自传里，格罗申迪克说当时的生活极不稳定，经常有周期性的扫荡，于是他与其他犹太人不得不经常在树林里一躲藏就是好几天。然而他还是坚持在当地镇上的学校接受一些教育。在17岁那年，他与母亲团聚。他断断续续在地方的一所大学学了一些课程，三年后，格罗申迪克的一位老师建议他去巴黎，向伟大的代数学家嘉当拜师。格罗申迪克听从了老师的建议，他的数学生涯也由此步入正轨。

格罗申迪克在自传中表示出对自己父母的敬重。如他的父亲一样，这个男人不随意改变自己的信念。1966年他获得菲尔兹奖，但却拒绝赴莫斯科参加国际数学大会领奖。1967年，他长途跋涉三周时间，来到越南北部，在森林里开设范畴论讲座。因为战争，河内的学生都被疏散到森林里以躲避美军的轰炸。

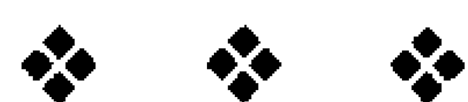
11年后，在他60岁时，格罗申迪克被瑞典皇家科学院授予克拉夫奖。这个奖项也被他拒绝了。（这一奖项有20万美元奖金，格罗申迪克似乎从来就不在乎金钱。）在他的说明信里，格罗申迪克指责了数学家和科学家阴暗的道德规范，不久后这封信就被刊登在法国《世界报》上。

格罗申迪克所做的这一切都是因为他信奉无政府主义，而不是受到了法国知识分子所标榜的反美主义思潮的影响。尽管去了越南，但格罗申迪克也不像很多有名的法国知识分子那样信仰共产主义。作为一名无政府主义者和一个对政治一无所知的人，格罗申迪克可能认为任何政治体系都是邪恶的，所有军队都只是屠杀的工具，所有有钱人都是穷人的

压迫者。这个时代的一些后现代主义的说教似乎也进入了他的脑子里。在他的自传中他说：

> 每一门科学，当我们**不是将它作为权力和统治的工具**，而是作为我们人类世代以来努力追求知识的冒险历程，那么，它就是一种和谐，从一个时期到另一个时期，或多或少，巨大而又丰富：在不同的时代和世纪中，对于依次出现的所有主题，它展现给我们微妙而精细的对应，仿佛来自虚空。

如果忽略上面这段话中后现代主义的粗体部分的词句，实际上这段行文非常漂亮。



再来说说他神圣而疯狂的一面。从IHES离职后，格罗申迪克到巴黎的法兰西学院教了两年书，但是他有一个令人不安的习惯，他不好好讲课而是替和平主义者呐喊。[9]法兰西学院几次与他沟通均无效，1973年，他在地中海沿岸、马赛以西的蒙彼利埃大学取得了一个职位。按照法国学术界的标准，这是超乎寻常的自我降级。大多数法国学术界人士要花几年时间筹划在巴黎得到一个职位，得到职位后，无论如何受尽折磨都要忍受而不能放弃。格罗申迪克对巴黎视而不见。世上没有什么事对他有意义。

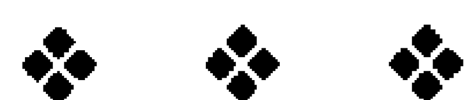
在蒙彼利埃的15年间，格罗申迪克写了他的自传《收获与播种》（从来没有出版，但其手稿却广为传播），还有其他数学和哲学著作和文章。1988年他60岁时退休。他学会了驾驶，当然水平遭透了。在日本，他成为年青人的偶像，而且受到佛教团体的僧侣们的欢迎。他变成了环保主义者，投身于环保运动，就这一问题和其他很多话题向权威们提出抗议。

1990年6月，他退休的两年后，格罗申迪克叫他的一位朋友保管他的所有数学论文。不久，也就是1991年初，他居然消失了。他的粉丝们最终在比利牛斯山脉的一个遥远的山村里找到了他，他在那里一直呆到今

天。有些报导说他皈依了佛教，也有报导说他责骂魔鬼和他的所有工作，以此来消磨时光。罗伊·利斯克来到格罗申迪克偏僻的隐居之处采访了他，并在2001年做了如下报导[10]。

> 尽管直接与他交流几乎是不可能的，但村子里的邻居们还是很照顾他。虽然据说他提出只靠喝蒲公英汤生存这样的想法，但是邻居们仍然尽量让他保持正常的日常饮食。这些邻居们也同巴黎和蒙彼利埃的好心人保持着联系，所以大家不必为他担心。

亚历山大·格罗申迪克在黄金时代所取得的数学成就现今依然很了不起，那些堪称理解它的人提起它的时候仍然带着几分敬畏和好奇。



人们常说一个国家的书面语言与人们平时讲话时的口语时近时远。英语在乔叟时代比较接近，而在18世纪早期的拉丁文学之黄金时代则相距甚远，而在我们自己的这个时代又更加相近。类似地，代数时而接近实际的科学世界，时而又远离它。

我已经说过，最早的代数来自于测量、计时和地形勘测等实际问题。(尽管没有什么实际意义，但是能够在我这本书的第一章和最后一章很自然提到土地勘测仍然是一件令人高兴的对称美。)丢番图和中世纪穆斯林数学家有时候脱离实际问题，为了他们自己的兴趣研究代数的话题，研究中已经带有抽象意味。这种研究态度被进一步带到文艺复兴时期和近代，当时的数学家对三次方程和四次方程的纯代数研究产生了极大的兴趣，特别是对一般解兴趣更大。

从大约1600年现代文字符号的发明开始，到18世纪后期攻克五次方程，新符号体系被广泛用于解决各个领域的实际问题：文化、军事工程、天文和航海、会计以及统计学的初级阶段。在这一时期，代数自美索不达米亚起源以来最为贴近实际问题的领域。

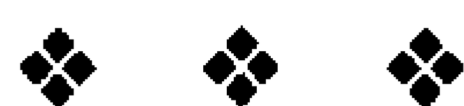
19世纪纯代数的发展异常丰富多彩，以致于这个学科超越了任何实

际应用，几乎只独居于那全无用途的领域。甚至当应用领域的人们从代数那里获得灵感时，他们也是做得漫不经心、囫囵吞枣。我在第8章已经提到吉布斯和赫维赛德对哈密尔顿珍爱的四元数的轻视。到了19世纪末，代数已经把科学抛到后面。在1893年，如果你向年轻的希尔伯特请教零点定理的实际应用，他也许会大笑起来。

尽管20世纪代数的抽象层次更高，但是此时可以看到这种缺口在某种程度上开始闭合。19世纪发现的所有新数学对象都找到了某些科学应用。1960年尤金·维格纳在他的一篇非常有影响的随笔中说这是奇迹，"在自然科学中数学的影响有些过了"。事实证明，这些纯推理的产物，这些群和矩阵，这些域和流形，是现实世界中某些真实事物或真实过程的化身。

代数的"过度影响"已经随处可见。例如，群在编码和加密理论中非常重要，矩阵现在是经济分析的基础，代数拓扑的一些概念出现在从发电到计算机芯片设计的各个领域。甚至范畴论也在计算机语言设计中发挥着神奇的功效，尽管我自己无法判断这一说法的价值。

然而，毫无疑问，维格纳的"过度影响说"给人印象最深的证明发生在现代物理学中，至少对于代数来说是这样。



20世纪物理学发生的两个伟大的革命当然是相对论和量子理论。它们都是建立在19世纪的纯代数概念基础之上的。

- 在狭义相对论中，利用洛伦兹变换，可以把一个参照系下的时间和空间的测定"转化"成另一个（以相对于第一个参照系匀速移动的）参照系下的时间和空间的测定。这些变换可以建模为特定四维空间坐标系中的旋转，换句话说，就是可以建模为一个李群。
- 在广义相对论中，因物质和能量的存在，这一四维时空被扭曲。为了对此给出恰当的解释，我们必须依靠张量演算，这是意大利代数几何学家根据哈密尔顿、黎曼和格拉斯曼的初期工作而发展

起来的。

- 1925年春天，当年轻的物理学家维尔纳·海森堡研究当一个原子从一个量子状态“跳到”另一个量子状态时放射出的辐射频率时，他觉得自己正在观察一个大的正方形数组，这个数组的第n列第m行上的数值是这个量子从状态m跳到状态n的概率。这种情况的推理需要他把这些数组乘起来，并给出这样做的合适的方法，但是当他尝试着做这样的乘法时，他发现数组是不可交换的。数组A乘以数组B得到一个结果，而数组B乘以数组A则得到完全不同的结果。究竟是怎么回事？好在海森堡是哥廷根大学的研究助理，所以他身边有希尔伯特和诺特来为他解释矩阵代数的原理。
- 20世纪60年代早期，物理学家发现了一种令人困惑不解的核粒子怪物——强子。加州理工学院的年轻物理学家马瑞·盖尔曼注意到，尽管强子的性质不遵从显然的线性模式，但是在另一个李群背景下还是有意义的，这个李群就是我们研究复数坐标的二维空间中的旋转时出现的群。对数据进行研究之后，盖尔曼发现原来的想法还是太肤浅了。复数坐标的三维空间中的等价李群能够解释得更清楚。然而，它需要还没有观察到的粒子的存在。盖尔曼发表了他的研究结果，实验员加大了强子对撞器的动力，适时观察到了他预测的粒子。[11]

现在，在21世纪初期，正在研究更奇怪更大胆的物质理论。如果没有哈密尔顿、格拉斯曼、柯西、西尔维斯特、希尔伯特和诺特的工作，这些物质理论都不能得到证实。其中最大胆的理论是要统一20世纪两个伟大发现——相对论和量子力学，从而又冒出了很多新名词：弦论、超对称弦论、M理论和圈量子引力。所有这一切都至少有一部分灵感来自于20世纪的代数或代数几何。

以卡拉比-丘流形（见图15-4）为例，它提供了弦论所需要的“消失”维度。根据弦论，这些是潜伏在时空中以普朗克长度（约为1.6×10^{-33}cm）

计数的极小区域中的六维空间。它们是由德国数学家埃里希·卡勒尔（1902—2000）首先想到的，虽然晚了几年（1932年到1933年发现的）。卡勒尔同扎里斯基一样来到罗马和意大利代数几何学家一起进行研究工作。

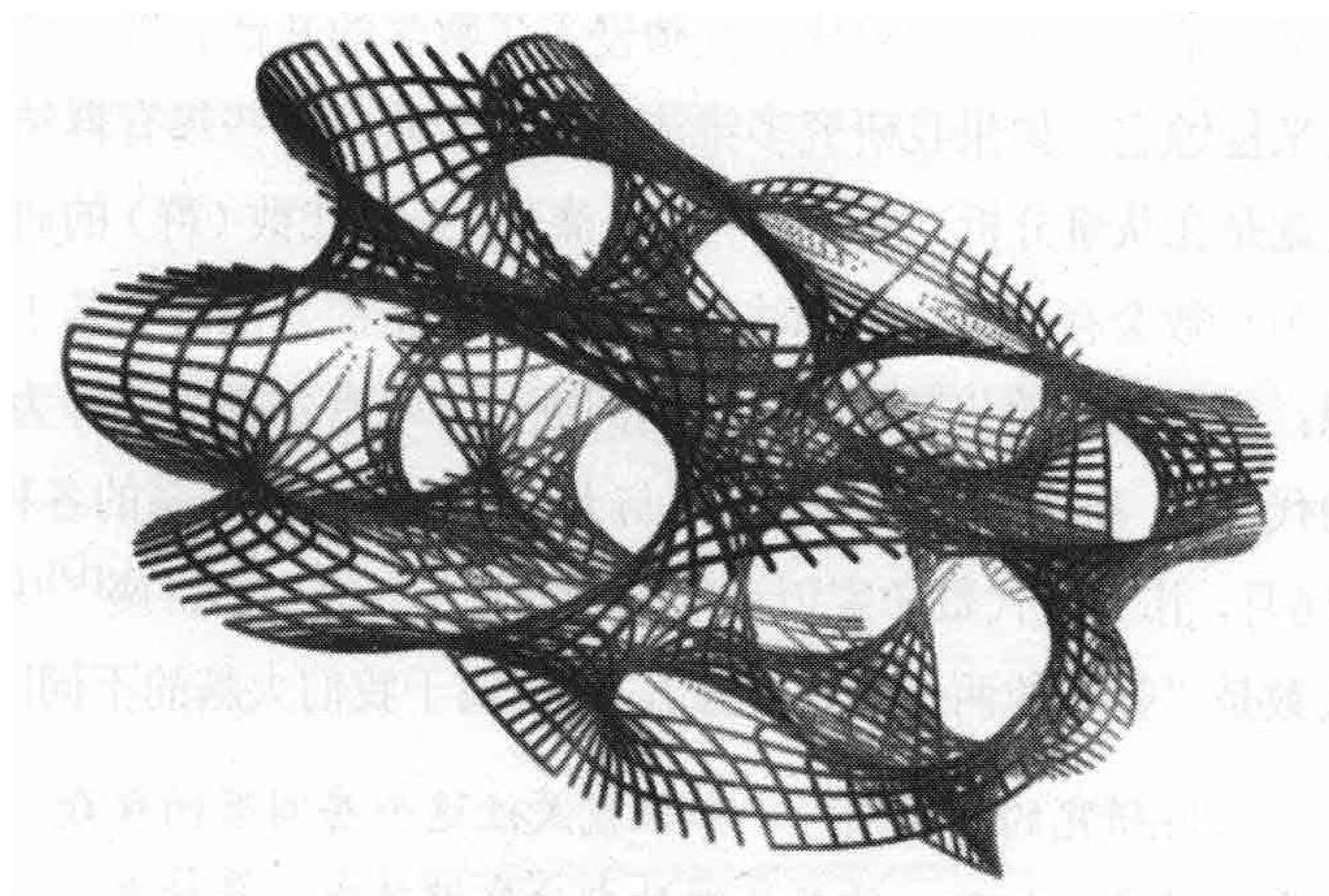
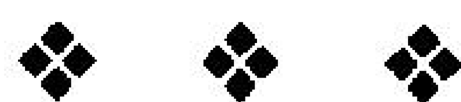


图15-4　卡拉比-丘流形

利用黎曼的某些思想，卡勒尔定义了具有特定的一般性和有趣性质的流形族[12]。例如，每个黎曼曲面都是一个卡勒尔流形。下一代美国数学家欧根尼奥·卡拉比确定了卡勒尔流形的一个子类，并推测说它们的弯曲应该有一种有趣的单纯性。

1977年，来自中国的年轻数学家丘成桐[13]证明了卡拉比猜测。这种类型的空间现在被称为卡拉比-丘流形。它们弯曲的单纯性，即某种“平滑”的性质，使得它们对这样的弦运动非常理想：根据弦论，这种运动在我们的各种仪器中是作为多种亚原子粒子以及包括引力在内的各种力的运动出现的。它们是六维空间的事实有点令人惊讶，但是这些“额外”的维度被“卷了起来”，在我们主导的宏观世界中不可见，就像从足够远的地方看一个粗粗的三维钢索时它像是一维的点一样。



因此，似乎有理由这样想，刻画20世纪代数特征的对高层次抽象的

追求也许会停止，或者至少暂时停止，而同时，代数学家们忙于回答物理学家提出的难题，而且要摆正诸如范畴论这样超抽象方法的地位。

也有这样的可能，作为数学的一个独立学科，代数也许无法存活下去。20世纪是一个统一的时代，代数侵入了数学的其他领域，这些领域也反过来侵蚀它。如果我研究多维流形上函数族，这些族有群结构，那么我究竟是在从事分析（函数）、拓扑（流形）还是代数（群）的研究呢？

认为代数会存活下去的想法，也是我所希望的想法，它基于这样一种思想：存在独特的代数思维。我们回顾一下哈密尔顿的“作为纯时间科学的代数”，和其他关于数学思维与人类思维活动的关系的各种思考。2000年6月，伟大的代数学家迈克尔·阿蒂亚在多伦多的讲座[14]中说，几何和代数是“数学的两个形式支柱”，它们属于我们大脑的不同区域。

> 几何研究的是空间……如果我关注这个房间里的观众，我能看到很多，仅用一秒或几微秒我就能够获取大量信息……另一方面，代数本质上与时间相关。无论你研究的是何种代数，一个接一个地进行一系列操作表明你需要时间。在静态的宇宙中，你无法想象代数，但是几何本质上是静态的。

迈克尔爵士说的“一系列操作”就是算法，而这一单词是（见第3章）给出代数这个词的那个人的名字的讹用版本。

据我所知，数学家埃里克·格伦沃尔德在2005年春天的《数学信使》杂志中把迈克尔爵士的思维推向了极致。格伦沃尔德在标题为“数学内外的进展与构思”一文中赞成二分思维，也就是一种阴-阳二进制方案，对此我概述如下。（有一些条目是我自己加上的，与格伦沃尔德无关。）

阴	阳
几何	代数
发现	发明
看	听

（续）

阴	阳
绘画	音乐
说明性的	描述性的
理论构建	问题求解
安全	冒险
空间模式	穿越时空的过程
牛顿	莱布尼茨[15]
庞加莱	希尔伯特
爱因斯坦[16]	马赫
设计	演化
社会主义	资本主义
柏拉图哲学的（数学观点）	“社会构架”
理论的（物理学）	实验的

当然，你可以整个晚上都玩这种智力游戏但得不出什么结论。（我真佩服自己有这么强的自控能力，居然没把“奥古斯丁派”和“伯拉纠派”放在这个列表中。）

然而，我认为迈克尔爵士和格伦沃尔德意识到了某种东西。今天，数学已经在最高层次上得到了完美的统一，一个传统领域（几何、数论）的概念轻易地融入到了另一个领域（代数、分析）。然而，仍然存在不同的思维风格、解决问题及获得新见解的方法。几年前我们就听到很多关于是否历史已经终结的谈话。我想不起来我们的博学者或哲学家关于这个大问题是否得出了什么结论，但是我确信至少代数的历史还没有结束。

注　解

[1] 数学家巴里·马佐尔教授同时也是一位经验丰富、文笔流畅的数学科普作家，他在2004年11月的《美国数学协会公告》上向非代数读者解释“原动力”这个概念。他的文章是这样开始的：“一个有限连通单纯复形X的代数拓扑能够被一维上同调刻画到什么程度呢？”我认为这篇文章已经把这一概念解释得再清楚不过了。

[2] 根据美国数学协会的分类码，这13个主题是：（06）秩、格、有序代数结构，（08）一般代数体系，（12）域论和多项式，（13）可交换环和代数，（14）代数几何，（15）线性和多重线性代数、矩阵理论，（16）结合环和代数，（17）非结合环和代数，（18）范畴论、同调代数，（19）K理论，（20）群论及其扩展，（22）拓扑群、李群，（55）代数拓扑。

[3] 麦克莱恩说，老伯克霍夫的偏见至少部分是由20世纪30年代大失业导致的朴素爱国主义促成的，对此我无法判断是否正确。麦克莱恩是这样说的："哈佛的乔治·伯克霍夫……认为我们还应该关注年轻的美国人，使得在哈佛给［欧洲］难民少留一些职位。"（摘自《更多数学人》（*More Mathematical People*)一书。）

[4] 斯旺教授对此有一段有趣的历史注释："同伦群是由爱德华·塞赤于1932年发现的，但是当他发现它们的特性主要是可交换的时，他认为它们没有意思，就撤回了论文。几年之后，维托尔德·胡尔维茨重新发现了它们。因此通常认为发现同伦群的荣誉应该属于维托尔德·胡尔维茨。"

[5] 这是一个与几何中到处可见的对偶的概念相关的过程。三维几何中经典的"柏拉图多面体"就展示了对偶。立方体（8个顶点，12条边，6个侧面）与八面体对偶（6个顶点，12条边，8个侧面），十二面体（20个顶点，30条边，12个侧面）与二十面体对偶（12个顶点，30条边，20个侧面），四面体（4个顶点，6条边，4个侧面）与自身对偶。顺便提一下，正是爱米·诺特指出了关注群的这些性质的好处。早期的研究者们已经用某种程度上不同的语言描述了同调群。

[6] "4单形体"是在考克斯特的著作《正规多面体》中用于描述这一对象的一个词。我没有在其他地方看到这个词，也不知道现在使用什么词，也不认为它会一直延用下去。当然，这个金属框架的4单形体已经被从四维投射到了二维，所以图15-2是非常不充分的。

[7] 关于术语"泛代数"有一段很有趣的历史，这要追溯到1898年的一本书《泛代数》。这本书的作者是阿弗烈·诺夫·怀海德，他是数学哲学家，与伯特兰·罗素齐名，他们合著了《数学原理》一书。爱米·诺特也使用这一术语。我只是偶尔提示性地使用这一术语，与怀海德、诺特或者其他任何人的使用不完全一致。

[8] 据我所知，范畴论在流行文化中只出现过一次，就是在2001年的电影《美丽心灵》中。有这样一幕，一个学生对约翰·纳什说："伽罗华扩张实际上与覆盖空间一样！"于是这个学生一边吃着三明治一边列数相同之处："……函子……两个范畴……"这其中的含意似乎是伽罗华扩张（参见关于域论的数学知识）和覆盖空间（一个拓扑概念）是两个范畴，可以通过一个函子把一个映入到另一个，

这是一个相当敏锐的见解。

[9] 艾琳·杰克逊引用了贾丝廷·巴姆拜讲述的这件事的一些内部信息，当时格罗申迪克正与贾丝廷生活在一起："他的数学学生都很认真，而且也很守纪律，都是一些勤奋的人……在反传统文化中，他遇到了一些整天听音乐虚度光阴的人。"

[10] 格罗申迪克传记网站：www.fermentmagazine.org/home5.html。

[11] 洛伦茨群和盖尔曼用于组织强子的那个群，技术上就是秩为3的酉群，它们都可以用矩阵来建模，尽管矩阵中的项是复数。

[12] 流形的精确定义是："一个黎曼流形允许对于有反次对称扭曲的度量联络做并联旋。"

[13] 丘成桐是菲尔兹奖和克拉夫奖获得者。他1949年出生在中国广东省，后随家人来到了香港，在那里接受了早期的数学教育。现在他是哈佛大学数学教授。

[14] 在《美国数学月刊》第108（7）期上以"20世纪的数学"为题发表。

[15] 这里的意思是，牛顿是绝对空间的代表，而莱布尼茨则更倾向于观察，就如古老的小调解释的那样：

空间

是阻止每一种事物在相同地方的东西

[16] 这是一个非常普遍的误解，认为爱因斯坦把绝对排除在物理之外，而把我们投入到一个相对世界之中。事实上，他没有做这种事。爱因斯坦是比牛顿更"绝对"。他排除的是绝对空间和绝对时间，取而代之的是绝对时空。关于现代物理的任何一本受欢迎的书籍都应该清晰阐述这一点。另外，爱因斯坦的亲密朋友科特·歌德尔是一位严格的柏拉图主义者：俩兄弟都是阴。

插图说明

没用特别提及的图片和只提及了文件来源的图片，是超出著作权保护期限的图片或者无法确定其所有权的图片。

没有人比作者对版权法更敏感了，而且我已经尽全力去征得当事人的同意，让他们授权我再现这些照片和图片。尽管确定这些当事人不是一件容易的事情。当某人在因特网上发现这些图片时，这件事就会更麻烦，因为可能没有明确的归属人。

如果任何个人或团体拥有本书中出现的一些图示资料的著作权，我只能要求他通过出版商与我取得联系，容我做适当的赔偿，对此我很高兴，而且会尽快赔偿。

诺伊格鲍尔：承蒙位于罗得岛州普罗维登斯的布朗大学约翰·海伊图书馆惠允。

海帕希亚：英国画家查尔斯·威廉·米契尔（1854—1903）所作，承蒙位于英国泰恩河畔纽卡斯尔市泰恩和怀尔美术馆的复制惠允。

卡尔达诺：取自于卡尔达诺的著作《伟大的艺术，代数法则》的卷首插图。我是从这本著作的麻省理工出版社的版本中复制的，该版本由理查德·维特默翻译并编辑的（剑桥，马萨诸塞州，1968年）。

韦达：取自于《弗兰索瓦·韦达：歌剧数学》的卷首插图，得到弗朗西斯·A·斯库特惠允。该书由位于德国希尔德斯海姆市的Georg Olms Verlag出版社出版，我所使用的版本为1970年纽约出版的。

笛卡儿：这是一位我不知道的艺术家的雕刻画像，取材于弗兰兹·哈尔斯1649年的画作，这幅画现在保存在巴黎的罗浮宫。

牛顿：托马斯·奥海姆·巴洛1868年的一个雕刻画像，取材于戈弗雷·内勒1689年的画作，现在这幅肖像画保存在伦敦的威尔康图书馆。

莱布尼茨：以佛罗伦萨的乌菲奇美术馆的一幅油画为素材所做的雕刻画像。

拉菲尼：取自《保罗·拉菲尼的歌剧数学》第一卷的卷首插图；该书译自1915年的意大利文版本。

柯西：这是J·罗勒在大约1840年所作的油画，得到位于法国香榭丽舍河畔马恩的法国道桥学院的惠允。

阿贝尔：取材自位于巴黎的Gauthier-Villars出版社1885年出版的C·A·皮叶克尼斯的著作*Niels-Henrik Abel:Tableau de Sa Vie et Son Action Scientifique*。

伽罗华："伽罗华15岁的肖像"，取材于位于巴黎的Tome XIII出版社1896年出版的*Annales de l'Ecole Normale Supérieure*一书第三卷。

西罗：得到挪威奥斯陆大学图书馆惠允。

若尔当：取自于位于巴黎的Gauthier-Villars & Cie Editeur-Imprimeur-Libraire 出版社1961年出版的*Oeuvres de Camille Jordan*一书的第一卷，J·狄多涅编辑。

哈密尔顿：莎拉·帕瑟的油画，得到都柏林皇家爱尔兰学院图书馆惠允。

格拉斯曼：取自于位于纽约的切尔西出版公司于1969年出版的*Hermann Grassmanns Gesammelte Mathematische und Physikalische Werke*，得到美国数学协会惠允。

黎曼：承蒙柏林的德国国家图书馆惠允。

艾勃特：承蒙伦敦市中学惠允。

普吕克：取材于位于德国莱比锡的Druck und Verlag von B.G.Teubner出版社1895年出版的*Julius Plückers Gesammelte Mathematische Abhandlungen*一书，由A·绍恩佛莱斯编著。

李：约阿希姆·弗洛姆的油画，承蒙挪威奥斯陆大学惠允。

克莱因，戴德金，希尔伯特，诺特：承蒙德国哥廷根大学图书馆惠允。

莱夫谢茨：承蒙纽约新泽西州普林斯顿大学图书馆稀有图书和特殊收藏系惠允。

扎里斯基，格罗申迪克：承蒙德国沃尔法赫数学研究所档案馆惠允。

麦克莱恩：得到芝加哥大学惠允。

卡拉比-丘的插图（图15-4）是巴黎综合理工学院数学应用中心的让·弗朗西斯科·柯隆纳绘制的。本书中的复制得到了他的许可。